



TESIS - SS142501

**PEMODELAN REGRESI SEMIPARAMETRIK  
MENGUNAKAN ESTIMATOR CAMPURAN  
*SPLINE TRUNCATED* DAN KERNEL**

HESIKUMALASARI  
NRP. 1314 201 205

Pembimbing:  
Prof. Dr. Drs. I Nyoman Budiantara, M.Si.  
Dr. Vita Ratnasari, S.Si., M.Si.

PROGRAM MAGISTER  
JURUSAN STATISTIKA  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
INSTITUT TEKNOLOGI SEPULUH NOPEMBER  
SURABAYA  
2016



TESIS - SS142501

# **SEMIPARAMETRIC REGRESSION MODEL USING COMBINED ESTIMATOR TRUNCATED SPLINE AND KERNEL**

HESIKUMALASARI  
NRP. 1314 201 205

Supervisor:  
Prof. Dr. Drs. I Nyoman Budiantara, M.Si.  
Dr. Vita Ratnasari, S.Si., M.Si.

MAGISTER PROGRAM  
DEPARTMENT OF STATISTICS  
FACULTY OF MATHEMATICS AND NATURAL SCIENCES  
SEPULUH NOPEMBER INSTITUTE OF TECHNOLOGY  
SURABAYA  
2016

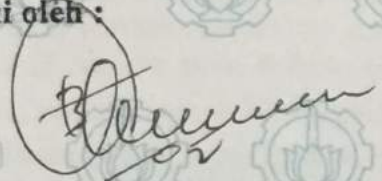
**PEMODELAN REGRESI SEMIPARAMETRIK MENGGUNAKAN  
ESTIMATOR CAMPURAN *SPLINE TRUNCATED* DAN KERNEL**

**Tesis disusun untuk memenuhi salah satu syarat memperoleh gelar  
Magister Sains (M.Si.)  
di  
Institut Teknologi Sepuluh Nopember  
Oleh :**

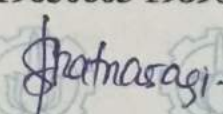
**HESIKUMALASARI  
NRP. 1314 201 205**

**Tanggal Ujian : 18 Juli 2016  
Periode Wisuda : September 2016**

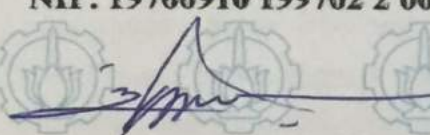
**Disetujui oleh :**

  
**1. Prof. Dr. Drs. I Nyoman Budiantara, M.Si.  
NIP. 19650603 198903 1 003**

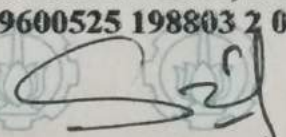
**(Pembimbing I)**

  
**2. Dr. Vita Ratnasari, S.Si., M.Si.  
NIP. 19700910 199702 2 001**

**(Pembimbing II)**

  
**3. Dr. Bra. Ismaini Zain, M.Si.  
NIP. 19600525 198803 2 001**

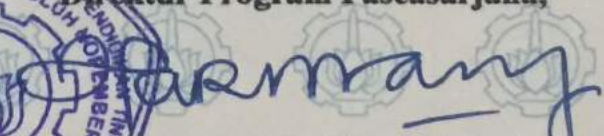
**(Penguji)**

  
**4. Dr. Santi Wulan Purnami, S.Si., M.Si.  
NIP. 19691212 199303 2 002**

**(Penguji)**



**Direktur Program Pascasarjana,**

  
**Prof. Ir. Djauhar Manfaat, M.Sc., Ph.D.  
NIP. 196012021987011001**

# PEMODELAN REGRESI SEMIPARAMETRIK MENGUNAKAN ESTIMATOR CAMPURAN *SPLINE* *TRUNCATED* DAN KERNEL

Nama : Hesikumalasari  
NRP : 1314201205  
Pembimbing : Prof. Dr. I Nyoman Budiantara, M.Si.  
Co-pembimbing : Dr. Vita Ratnasari, M.Si.

## ABSTRAK

Variabel respon pada analisis regresi dapat memiliki hubungan linear dengan salah satu variabel prediktor, namun dengan variabel prediktor yang lain tidak diketahui bentuk pola hubungannya. Pola data seperti ini dapat dimodelkan dengan model regresi semiparametrik. Diberikan data berpasangan  $(x_i, t_i, z_i, y_i)$  dan diasumsikan hubungan antar variabel prediktor dengan variabel respon mengikuti model regresi semiparametrik  $y_i = \mu(x_i, t_i, z_i) + \varepsilon_i$ . Kurva  $\mu(x_i, t_i, z_i)$  dapat ditulis dalam bentuk  $\mu(x_i, t_i, z_i) = f(x_i) + g(t_i) + h(z_i)$ . Kurva regresi  $f(x_i)$  dihampiri dengan parametrik linear,  $g(t_i)$  dihampiri dengan fungsi nonparametrik *spline truncated* dan  $h(z_i)$  dihampiri dengan fungsi nonparametrik kernel. Tujuan dari penelitian ini adalah memperoleh bentuk estimator dalam regresi semiparametrik dengan menggunakan estimator campuran *spline truncated* dan kernel. Berdasarkan hasil kajian diperoleh estimator kurva regresi parametrik linear adalah  $\hat{f}(x) = \mathbf{C}(\tilde{k}, \tilde{\alpha})\tilde{y}$ , estimator kurva regresi *spline truncated* adalah  $\hat{g}(t) = \mathbf{K}(\tilde{k}, \tilde{\alpha})\tilde{y}$  dan estimator kurva regresi kernel adalah  $\hat{h}(z) = \mathbf{D}(\tilde{\alpha})\tilde{y}$ . Selanjutnya diperoleh estimator kurva regresi semiparametrik campuran *spline truncated* dan kernel  $\hat{\mu}(x, t, z) = \mathbf{M}(\tilde{k}, \tilde{\alpha})\tilde{y}$ , dengan  $\mathbf{M}(\tilde{k}, \tilde{\alpha}) = \mathbf{C}(\tilde{k}, \tilde{\alpha}) + \mathbf{K}(\tilde{k}, \tilde{\alpha}) + \mathbf{D}(\tilde{\alpha})$ . Estimator campuran ini tergantung pada titik knot dan parameter *bandwidth*. Estimator terbaik diperoleh dengan cara meminimumkan fungsi *Generalized Cross Validation*. Model yang diperoleh kemudian diterapkan pada data produksi padi di Jawa Tengah. Model yang diperoleh adalah model estimator campuran yang menghasilkan  $R^2$  sebesar 91,91%.

*Kata Kunci:* Regresi Semiparametrik, *Spline Truncated*, Kernel, Estimator Campuran.

*Halaman ini sengaja dikosongkan*

# SEMIPARAMETRIC REGRESSION MODEL USING COMBINED ESTIMATOR TRUNCATED SPLINE AND KERNEL

Nama : Hesikumalasari  
NRP : 1314201205  
Pembimbing : Prof. Dr. I Nyoman Budiantara, M.Si.  
Co-pembimbing : Dr. Vita Ratnasari, M.Si.

## ABSTRACT

The response variable of the regression analysis has a linear relationship with one of the variable predictors inspite of the unknown relationship pattern with the other variables. Consequently, it can be approached by using semiparametric regression model. As given pair data  $(x_i, t_i, z_i, y_i)$  and assumed the relationship between predictor variable and response variable are approached by using semiparametric regression  $y_i = \mu(x_i, t_i, z_i) + \varepsilon_i$ , Combined curve  $\mu(x_i, t_i, z_i)$  can be written as  $\mu(x_i, t_i) = f(x_i) + g(t_i) + h(z_i)$ . Regression curve  $f(x_i)$  is investigated by using parametric linear, regression curve  $g(t_i)$  is investigated by using truncated spline function and  $h(z_i)$  is approached by using kernel function. The objective of this research is to obtain the pattern of combined estimator in semiparametric regression by using the estimator of a combination of truncated spline and kernel. Based on result obtained the estimator of linear parametric curve is  $\hat{f}(x) = \mathbf{C}(\tilde{k}, \tilde{\alpha})\tilde{y}$ , the estimator of spline truncated curve is  $\hat{g}(t) = \mathbf{K}(\tilde{k}, \tilde{\alpha})\tilde{y}$ , and the estimator of kernel curve is  $\hat{h}(z) = \mathbf{D}(\tilde{\alpha})\tilde{y}$ . Then the estimator of combined of spline truncated and kernel semiparametric curve is  $\hat{\mu}(x, t, z) = \mathbf{M}(\tilde{k}, \tilde{\alpha})\tilde{y}$ , where  $\mathbf{M}(\tilde{k}, \tilde{\alpha}) = \mathbf{C}(\tilde{k}, \tilde{\alpha}) + \mathbf{K}(\tilde{k}, \tilde{\alpha}) + \mathbf{D}(\tilde{\alpha})$ . The estimator of this combination depends on knot points and badwith parameter. Best estimator are resulted by minimizing the Generalized Cross Validation function. The combination of truncated spline and kernel for Semiparametric regression model which have been obtained will be applied on East Java's rice production data. Model that obtained is combined estimator model with  $R^2$  is 91.91%.

*Keywords* : Semiparametric Regression, Spline Truncated, Kernel, Combine Estimator

*Halaman ini sengaja dikosongkan*

## DAFTAR ISI

	Halaman
<b>HALAMAN JUDUL</b>	
<b>HALAMAN PENGESAHAN</b> .....	<b>i</b>
<b>ABSTRAK</b> .....	<b>iii</b>
<b>ABSTRACT</b> .....	<b>v</b>
<b>KATA PENGANTAR</b> .....	<b>vii</b>
<b>DAFTAR ISI</b> .....	<b>ix</b>
<b>DAFTAR GAMBAR</b> .....	<b>xi</b>
<b>DAFTAR TABEL</b> .....	<b>xiii</b>
<b>DAFTAR LAMPIRAN</b> .....	<b>xv</b>
<b>BAB I PENDAHULUAN</b> .....	<b>1</b>
1.1 Latar Belakang .....	1
1.2 Rumusan Masalah .....	4
1.3 Tujuan Penelitian .....	5
1.4 Manfaat Penelitian .....	5
1.5 Batasan Masalah .....	5
<b>BAB II TINJAUAN PUSTAKA</b> .....	<b>7</b>
2.1 Analisis Regresi .....	7
2.2 Regresi Parametrik, Nonparametrik dan Semiparametrik .....	7
2.3 Regresi Parametrik Linear .....	9
2.4 Estimator Spline .....	10
2.5 Estimator Kernel .....	13
2.6 <i>Generalized Cross Validation</i> (GCV) .....	15
2.6 Koefisien Determinasi .....	15
2.7 Produksi Padi .....	16
<b>BAB III METODOLOGI PENELITIAN</b> .....	<b>21</b>
3.1 Sumber Data .....	21
3.2 Variabel Penelitian dan Struktur Data .....	21
3.3 Definisi Operasional .....	22



3.4 Metode Penelitian .....	22
<b>BAB IV HASIL DAN PEMBAHASAN .....</b>	<b>27</b>
4.1 Model Regresi Semiparametrik Campuran .....	27
4.1.1 Model Regresi Semiparametrik Campuran Univariabel .....	27
4.1.2 Model Regresi Semiparametrik Campuran Multivariabel....	31
4.2 Estimasi Kurva Regresi Semiparametrik Campuran .....	34
4.3 Aplikasi pada Data Produksi Padi .....	39
4.3.1 Analisis Deskriptif Data .....	39
4.3.2 Model Regresi Semiparametrik Campuran .....	45
4.3.2.1 Model dengan 1 Titik Knot .....	46
4.3.2.2 Model dengan 2 Titik Knot .....	47
4.3.2.3 Model dengan 3 Titik Knot .....	48
4.3.3 Interpretasi Model Regresi Semiparametrik Campuran .....	59
<b>BAB V KESIMPULAN DAN SARAN .....</b>	<b>63</b>
5.1 Kesimpulan .....	63
5.2 Saran .....	64
<b>DAFTAR PUSTAKA .....</b>	<b>65</b>
<b>LAMPIRAN .....</b>	<b>69</b>

## DAFTAR TABEL

	Halaman
Tabel 3.1	Variabel Penelitian 21
Tabel 3.2	Struktur Data Penelitian 21
Tabel 4.1	Deskriptif Data 40
Tabel 4.2	GCV dari Model dengan 1 Spline dan 3 Kernel 43
Tabel 4.3	GCV dari Model dengan 2 Spline dan 2 Kernel 43
Tabel 4.4	GCV dari Model dengan 3 Spline dan 1 Kernel 44
Tabel 4.5	Komponen Parametrik dan Nonparametrik 44
Tabel 4.6	Perbandingan Nilai GCV Model dengan Komponen Spline 1 Titik Knot 46
Tabel 4.7	Estimasi Parameter Komponen Spline Truncated dengan 1 Titik Knot 46
Tabel 4.8	Perbandingan Nilai GCV Model dengan Komponen Spline 2 Titik Knot 47
Tabel 4.9	Estimasi Parameter Komponen Spline Truncated dengan 2 Titik Knot 48
Tabel 4.10	Perbandingan Nilai GCV Model dengan Komponen Spline 3 Titik Knot 49
Tabel 4.11	Estimasi Parameter Komponen Spline Truncated dengan 3 Titik Knot 50
Tabel 4.12	Perbandingan GCV untuk model dengan Komponen Spline Truncated 1, 2 dan 3 Knot 51
Tabel 4.13	Perbandingan Nilai $R^2$ 58

*Halaman ini sengaja dikosongkan*

## DAFTAR GAMBAR

	Halaman
Gambar 3.1 Diagram Alir Langkah Analisis Pertama	25
Gambar 3.2 Diagram Alir Langkah Analisis Kedua	26
Gambar 4.1 Scatterplot antara Jumlah Produksi Padi dengan Luas Panen	40
Gambar 4.2 Scatterplot antara Jumlah Produksi Padi dengan Penggunaan Pupuk	40
Gambar 4.3 Scatterplot antara Jumlah Produksi Padi dengan Penggunaan Pestisida	41
Gambar 4.4 Scatterplot antara Jumlah Produksi Padi dengan Penggunaan Benih	41
Gambar 4.5 Scatterplot antara Jumlah Produksi Padi dengan Penggunaan Tenaga Kerja	42
Gambar 4.6 Plot data $y$ dan $\hat{y}$ Data ke-1 Sampai Data ke-50	51
Gambar 4.7 Plot data $y$ dan $\hat{y}$ Data ke-51 Sampai Data ke-100	52
Gambar 4.8 Plot data $y$ dan $\hat{y}$ Data ke-101 Sampai Data ke-150	52
Gambar 4.9 Plot data $y$ dan $\hat{y}$ Data ke-151 Sampai Data ke-200	53
Gambar 4.10 Plot data $y$ dan $\hat{y}$ Data ke-201 Sampai Data ke-250	53
Gambar 4.11 Plot data $y$ dan $\hat{y}$ Data ke-251 Sampai Data ke-300	54
Gambar 4.12 Plot data $y$ dan $\hat{y}$ Data ke-301 Sampai Data ke-350	54
Gambar 4.13 Plot data $y$ dan $\hat{y}$ Data ke-351 Sampai Data ke-400	55
Gambar 4.14 Plot data $y$ dan $\hat{y}$ Data ke-401 Sampai Data ke-450	55
Gambar 4.15 Plot data $y$ dan $\hat{y}$ Data ke-451 Sampai Data ke-500	56
Gambar 4.16 Plot data $y$ dan $\hat{y}$ Data ke-501 Sampai Data ke-550	56
Gambar 4.17 Plot data $y$ dan $\hat{y}$ Data ke-551 Sampai Data ke-600	57
Gambar 4.18 Plot data $y$ dan $\hat{y}$ Data ke-601 Sampai Data ke-650	57
Gambar 4.19 Plot data $y$ dan $\hat{y}$ Data ke-651 Sampai Data ke-691	58

*Halaman ini sengaja dikosongkan*

## DAFTAR LAMPIRAN

	Halaman
Lampiran 1    Data yang digunakan	69
Lampiran 2    Syntax R Model Semiparametrik Campuran	70

*Halaman ini sengaja dikosongkan*

# **BAB I**

## **PENDAHULUAN**

### **1. 1 Latar Belakang**

Analisis regresi merupakan salah satu metode Statistika yang digunakan untuk mengestimasi pola hubungan antara variabel prediktor ( $X$ ) dengan variabel respon ( $Y$ ). Tujuan utama dalam analisis regresi adalah mencari bentuk estimasi kurva regresi. Pendekatan yang biasa digunakan untuk mengestimasi kurva regresi yaitu pendekatan regresi parametrik dan pendekatan regresi nonparametrik. Regresi parametrik dilakukan apabila bentuk kurva regresinya diketahui. Sedangkan regresi nonparametrik dilakukan jika bentuk kurva regresinya tidak diketahui.

Dalam analisis regresi terdapat beberapa kasus dimana fungsi regresinya memuat dua komponen yaitu komponen parametrik dan komponen nonparametrik. Model regresi yang memuat dua komponen tersebut dinamakan model regresi semiparametrik atau biasa disebut model linear parsial. Untuk mengestimasi fungsi regresi jenis ini dilakukan penggabungan dua pendekatan yaitu pendekatan parametrik dan pendekatan nonparametrik. Model regresi semiparametrik mulai berkembang sejak Wahba pada tahun 1990 mempublikasikan suatu model hubungan semiparametrik dari penggunaan listrik pada negara bagian di Amerika Serikat, dimana pendapatan dan harga listrik mempunyai pola parametrik, sedangkan temperatur berpola nonparametrik. Penelitian sebelumnya mengenai regresi semiparametrik telah dilakukan oleh Srinadi (2002), Asmin (2009), Amelia (2014), dan lainnya.

Pendekatan regresi parametrik untuk mengestimasi kurva regresi diasumsikan bentuk kurvanya diketahui seperti linier, kuadratik, kubik, eksponensial dan lain sebagainya. Sementara dalam permasalahan pola hubungan yang bentuk kurvanya tidak diketahui atau tidak mempunyai bentuk pola hubungan tertentu digunakan pendekatan regresi nonparametrik (Härdle, 1990). Pendekatan regresi nonparametrik tidak mensyaratkan variabel yang diteliti harus mempunyai kurva regresi dengan pola tertentu atau bentuk kurvanya diketahui. Oleh karenanya, dengan menggunakan pendekatan ini, akan lebih fleksibel. Pada



prosedur regresi nonparamterik, data akan mencari sendiri bentuk kurva regresinya tanpa dipengaruhi oleh subjektivitas peneliti (Eubank, 1988). Pendekatan model regresi nonparametrik mengasumsikan pola kurva tidak diketahui dan termuat di dalam suatu fungsi tertentu, seperti ruang *Sobolev*, ruang fungsi-fungsi kontinu  $C(0,\pi)$ , ruang *Hilbert*, ruang *Entropi*, atau ruang-ruang lainnya. Pengambilan asumsi ini tergantung pada sifat *smooth* yang dimiliki oleh fungsi yang polanya tidak diketahui tersebut (Budiantara, 2000). Selanjutnya estimasi fungsi regresi nonparametrik dilakukan berdasarkan data pengamatan dengan menggunakan teknik *smoothing*. Ada beberapa teknik estimasi dalam regresi nonparametrik antara lain estimator histogram, *Spline*, kernel, polinomial lokal, *deret fourier*, dan *wavelet* (Eubank, 1999).

*Spline* merupakan salah satu teknik estimasi regresi nonparametrik yang pertama kali dikembangkan oleh Whittaker pada tahun 1923 (Hardle, 1990). *Spline* merupakan salah satu model yang mempunyai interpretasi statistik dan interpretasi visual sangat khusus dan sangat baik (Eubank, 1988). *Spline* diperoleh berdasarkan optimasi yang merupakan perluasan dari optimasi-optimasi yang ada pada regresi parametrik. Sementara itu, pendekatan regresi nonparametrik yang lain, tidak memiliki tata cara optimasi seperti dalam *Spline*. Dalam pendekatan *Spline* terdapat optimasi *Penalized Least Square* (PLS) yang diperoleh dari hasil generalisasi metode *Least Square* (LS) yang ada pada regresi parametrik. Di samping itu terdapat pula optimasi *Penalized Likelihood* (PL) yang merupakan hasil generalisasi metode *Maksimum Likelihood* (ML) serta optimasi *Penalized Log Likelihood* (PLL) yang merupakan hasil generalisasi metode *Log Likelihood* (LL) dalam regresi parametrik (Budiantara, 2006). Model *Spline* memiliki kemampuan yang sangat baik dalam mengakomodir data yang mempunyai perilaku berubah-ubah pada sub-sub interval tertentu (Cox dan O'Sullivan, 1996). Penelitian sebelumnya mengenai regresi nonparametrik *Spline* telah dilakukan oleh Wang (1985), Wahba (1990), Eubank(1999), Weinert (2009), Hazelton dan Turlach (2011), Nafi (2010), Adyana (2011), Samsodin (2011), dan lainnya.

Salah satu estimator lain selain *Spline* yang sering digunakan dalam regresi nonparametrik yaitu estimator kernel. Estimator kernel diperkenalkan pertama kali oleh Rosenblatt (1956) dan Parzen (1962) yang disebut estimator

densitas kernel Rosenblatt-Parzen (Hardle, 1990). Kelebihan estimator kernel adalah fleksibel, bentuk matematisnya mudah, dan dapat mencapai tingkat kekonvergenan yang relatif cepat (Budiantara dan Mulianah, 2007). Penelitian sebelumnya mengenai regresi nonparametrik *Spline* telah dilakukan oleh Muller (1984), William (1989), Okumura dan Naito (2006), Kulasekera dan Gallagher (2011), Eichner dan Stute (2012), Chamidah (2002), Zulfikar (2005), Susianto (2008), dan lainnya.

Model-model regresi nonparametrik maupun semiparametrik yang dikembangkan oleh para peneliti selama ini pada umumnya menggunakan jenis metode estimasi yang sama untuk sebagian bahkan semua prediktornya. Hal ini disebabkan oleh anggapan bahwa pola dari masing-masing prediktor dianggap mempunyai pola yang sama sehingga peneliti menggunakan hanya satu bentuk estimator model untuk semua variabel prediktor. Sementara itu, pada kenyataannya sering dijumpai kasus-kasus dengan pola data yang berbeda dari masing-masing variabel prediktor. Oleh karena itu, untuk mengatasi masalah tersebut beberapa peneliti telah mengembangkan estimator kurva regresi campuran nonparametrik dimana masing-masing pola data dalam model regresi nonparametrik dihipotesiskan dengan estimator kurva yang sesuai dengan pola data. Penelitian sebelumnya yang menggunakan estimator campuran pernah dilakukan oleh Sudiarsa (2015) yaitu melibatkan estimator deret fourier dan *Spline truncated*. Sedangkan penelitian yang dilakukan Budiantara (2015), Rory (2016) dan Purnomo (2016) telah melibatkan kernel dan *Spline truncated*. Penelitian-penelitian tersebut melibatkan estimator campuran pada regresi nonparametrik. Sementara itu, belum ada peneliti yang melibatkan estimator campuran pada regresi semiparametrik. Oleh karena itu, pada penelitian ini akan digunakan model campuran yaitu *Spline truncated* dan kernel pada regresi semiparametrik.

Tanaman padi adalah salah satu komoditas pangan utama karena merupakan tanaman penghasil beras yang menjadi makanan pokok bagi mayoritas penduduk di Indonesia. Kebutuhan padi di Indonesia terus meningkat dari tahun ke tahun dikarenakan jumlah penduduk yang terus meningkat. Peningkatan jumlah penduduk secara otomatis akan menambah konsumsi beras dalam negeri. Oleh karena itu produksi padi perlu ditingkatkan agar dapat memenuhi kebutuhan

tersebut. Berdasarkan data produksi padi beberapa tahun terakhir ini, produksi padi di Indonesia mengalami ketidakstabilan. Pada tahun 2009 produksi padi sebesar 64,40 juta ton gabah kering giling (GKG), kemudian pada tahun 2011 meningkat menjadi 65,76 juta ton. Namun demikian pada tahun 2014 mengalami penurunan sebanyak 0,84 juta ton dibandingkan tahun 2013 yang mencapai 71,28 juta ton.

Pusat produksi tanaman padi di Indonesia terdapat di Pulau Jawa. Pulau Jawa memiliki rata-rata produksi komoditas padi lebih tinggi dibandingkan di luar Pulau Jawa. Produksi padi Pulau Jawa pada tahun 2012 dan 2013 sebesar 36,52 juta ton dan 37,49 juta ton. Angka ini lebih tinggi dibanding daerah di luar Pulau Jawa dengan produksi hanya sebesar 32,2 juta ton dan 33,78 juta ton. Produksi padi terbanyak ini dihasilkan oleh Jawa Timur dan Jawa Barat, sementara Jawa Tengah masih berada dibawah angka produksi dua provinsi tersebut. Bahkan dalam kurun waktu antara tahun 2010-2014, provinsi Jawa Tengah mengalami dua kali penurunan hasil produksi padi, yakni pada tahun 2011 dan 2014. Penelitian sebelumnya tentang produksi padi pernah dilakukan oleh Susianto (2008) dan Asriani (2013), dimana pada penelitian-penelitian tersebut diketahui bahwa luas panen berhubungan linear terhadap produksi padi dan variabel lainnya berhubungan nonparametrik terhadap produksi padi. Susianto (2008) mendekati kurva regresi untuk variabel nonparametriknya menggunakan kernel dan Asriani (2013) menggunakan deret fourier. Sementara itu, dengan mempertimbangkan perbedaan pola data yang dapat dilihat berdasarkan *scatter plot* dari masing-masing variabel prediktor terhadap variabel respon, maka bisa didekati dengan estimator yang berbeda yang sesuai dengan pola data. Sehingga, berdasarkan perumusan masalah di atas, maka tujuan berikutnya dari penelitian ini adalah untuk menganalisis faktor-faktor yang berpengaruh terhadap produksi padi di Provinsi Jawa Tengah menggunakan regresi semiparametrik campuran *Spline truncated* dan kernel.

## **1.2 Rumusan Masalah**

Berdasarkan uraian latar belakang di atas, maka masalah dalam penelitian ini dapat dirumuskan sebagai berikut :

1. Bagaimana bentuk estimasi kurva regresi semiparametrik campuran *Spline truncated* dan kernel?
2. Bagaimana menerapkan estimasi campuran *Spline truncated* dan kernel untuk memodelkan data produksi padi di Jawa Tengah.

### 1.3 Tujuan Penelitian

Berdasarkan rumusan masalah di atas maka tujuan yang ingin dicapai adalah

1. Mengkaji bentuk estimasi kurva regresi semiparametrik campuran *Spline truncated* dan kernel.
2. Mengaplikasikan/menerapkan estimasi campuran *Spline truncated* dan kernel untuk memodelkan data produksi padi di Jawa Tengah.

### 1.4 Manfaat Penelitian

Hasil dari penelitian ini diharapkan dapat memberikan manfaat, antara lain :

1. Memberikan wawasan baru mengenai pemodelan, khususnya kurva regresi semiparametrik campuran *Spline truncated* dan kernel.
2. Mengetahui penerapan hasil estimasi yang diperoleh untuk memodelkan data produksi padi di Jawa Tengah.

### 1.5 Batasan Masalah

Mengacu pada permasalahan di atas, ruang lingkup dalam penelitian ini dibatasi pada beberapa hal, antara lain :

1. Data yang digunakan adalah data produksi padi di Jawa Tengah Tahun 2008.
2. Fungsi *Spline* yang digunakan dalam estimator campuran adalah *Spline truncated* linear.
3. Estimator kernel yang digunakan dalam estimator campuran adalah estimator kernel Nadaraya-Watson dan fungsi Kernel yang digunakan adalah Kernel Gaussian.
4. Pemilihan titik knot dan parameter *bandwidth* optimal menggunakan metode GCV.
5. Dalam aplikasi model digunakan *Spline truncated* linear dengan satu, dua dan tiga titik knot.

*Halaman ini sengaja dikosongkan*



## BAB II

### TINJAUAN PUSTAKA

#### 2.1 Analisis Regresi

Analisis regresi merupakan sebuah metode Statistika yang memberikan penjelasan tentang pola hubungan (model) antara dua variabel atau lebih. Bentuk pola hubungan kedua variabel tersebut dapat diidentifikasi berdasarkan suatu *scatter plot* atau pengalaman masa lalu yang memuat informasi tentang kurva regresi. *Scatter plot* merupakan alat bantu yang bisa menjelaskan apakah suatu kurva regresi bentuk polanya diketahui atau tidak diketahui. *Scatter plot* akan membantu untuk memberikan informasi apakah suatu fungsi dari kurva regresi mengikuti pola linear, kuadratik, polinomial, atau bahkan tidak mengikuti suatu pola tertentu. Misalkan diberikan data  $(x_i, y_i), i = 1, 2, \dots, n$  maka model regresinya dapat ditulis :

$$y_i = f(x_i) + \varepsilon_i, \quad i = 1, 2, \dots, n \quad (2.1)$$

dengan  $y_i$  adalah respon ke- $i$ ,  $f(x_i)$  adalah kurva regresi dan  $\varepsilon_i$  adalah *error* dimana  $\varepsilon_i \sim N(0, \sigma^2)$ .

#### 2.2 Regresi Parametrik, Nonparametrik dan Semiparametrik

Regresi parametrik merupakan metode yang digunakan untuk mengetahui pola hubungan antara variabel respon dengan variabel prediktor dengan asumsi bentuk kurva regresi diketahui. Apabila dalam analisis regresi bentuk kurva regresi  $f(x_i)$  pada persamaan (2.1) diketahui maka pendekatan model regresi tersebut dinamakan model regresi parametrik. Hubungan antara variabel respon dan variabel prediktor dalam model dapat terjadi dengan fungsi linier maupun nonlinier dalam parameter. Ada beberapa jenis regresi parametrik antara lain regresi linear, kuadratik, kubik dan lain sebagainya.

Sementara itu, regresi nonparametrik mulai dikenal sekitar abad XIX, tepatnya pada tahun 1857 (Hardle, 1990). Regresi nonparametrik adalah regresi yang digunakan jika hubungan antara variabel prediktor dengan variabel respon tidak diketahui bentuk kurvanya. Kurva  $f(x_i)$  pada persamaan (2.1) diasumsikan

tidak diketahui polanya maka digunakan pendekatan nonparametrik. Kurva fungsi diasumsikan termuat dalam ruang fungsi tertentu (Eubank, 1999). Pada analisis regresi nonparametrik tidak terikat pada bentuk kurva regresi yang diketahui. Kurva regresi hanya diasumsikan *smooth* (halus) dalam arti termuat di dalam suatu ruang fungsi tertentu. Pendekatan regresi nonparametrik memiliki fleksibilitas yang tinggi, karena data diharapkan mencari sendiri bentuk estimasi kurva regresinya tanpa dipengaruhi oleh faktor subyektivitas peneliti (Eubank, 1988). Ada beberapa teknik *smoothing* yang digunakan dalam regresi nonparametrik, diantaranya adalah estimator kernel, estimator *Spline*, estimator *penalized Spline*, estimator deret fourier, estimator wavelet, estimator deret ortogonal dan estimator polinomial lokal (Chamidah, 2008).

Sementara itu, regresi semiparametrik diperoleh dari gabungan antara komponen parametrik dan komponen nonparametrik (Budiantara, 2009). Dalam beberapa kasus, variabel respon dapat memiliki hubungan linear dengan salah satu variabel prediktor, tetapi dengan variabel prediktor yang lain tidak diketahui bentuk pola hubungannya. Variabel yang memiliki pola data yang dapat diketahui, atau terdapat informasi mengenai pola data digolongkan pada komponen parametrik. Sedangkan variabel yang tidak diketahui pola datanya digolongkan pada komponen nonparametrik. Diberikan pasangan data  $(x_i, t_i, y_i)$ , dimana  $i = 1, 2, \dots, n$  maka model semiparametrik dirumuskan sebagai berikut :

$$y_i = \mu(x_i, t_i) + \varepsilon_i \quad (2.2)$$

dengan kurva diasumsikan bersifat additive, yaitu :

$$\mu(x_i, t_i) = f(x_i) + g(t_i) \quad (2.3)$$

Sehingga persamaan (2.2) dapat ditulis menjadi :

$$y_i = f(x_i) + g(t_i) + \varepsilon_i \quad (2.4)$$

$y_i$  adalah variabel respon ke- $i$ ,  $f(x_i)$  adalah komponen parametrik,  $g(t_i)$  adalah fungsi regresi yang tidak diketahui atau merupakan komponen nonparametrik, dan  $\varepsilon_i$  adalah *error* dimana  $\varepsilon_i \sim N(0, \sigma^2)$ .



### 2.3 Regresi Parametrik Linear

Diberikan data berpasangan  $(x_i, y_i)$ ,  $i = 1, 2, \dots, n$ , dimana pola hubungannya dapat dinyatakan dalam model regresi  $y_i = f(x_i) + \varepsilon_i$ . Kurva regresi  $f(x_i)$  dihampiri dengan kurva regresi parametrik linear dimana :

$$f(x_i) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} \quad (2.5)$$

Sehingga model regresi linear berganda dengan  $p$  variabel prediktor dapat disajikan dalam bentuk :

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} + \varepsilon_i, \quad i = 1, 2, \dots, n \quad (2.6)$$

dimana  $y_i$  adalah variabel respon,  $\beta_0$  adalah konstanta,  $\beta_1, \beta_2, \dots, \beta_p$  adalah koefisien variabel prediktor, dan  $\varepsilon_i$  adalah *error* dimana  $\varepsilon_i \sim N(0, \sigma^2)$ .

Model regresi (2.2) dapat ditulis sebagai berikut :

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} \beta_0 + \beta_1 x_{11} + \beta_2 x_{12} + \dots + \beta_p x_{1p} \\ \beta_0 + \beta_1 x_{21} + \beta_2 x_{22} + \dots + \beta_p x_{2p} \\ \vdots \\ \beta_0 + \beta_1 x_{n1} + \beta_2 x_{n2} + \dots + \beta_p x_{np} \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

dapat ditulis dalam bentuk matriks :

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1p} \\ 1 & x_{21} & x_{22} & \dots & x_{2p} \\ 1 & x_{31} & x_{32} & \dots & x_{3p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{np} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

atau dapat ditulis dalam notasi matriks :

$$\tilde{y} = \mathbf{X}\tilde{\beta} + \tilde{\varepsilon} \quad (2.7)$$

dengan  $\tilde{y}$  adalah vektor dari variabel respon berukuran  $n \times 1$ ,  $\mathbf{X}$  merupakan matriks berukuran  $n \times (p + 1)$  dan  $\tilde{\beta}$  adalah vektor parameter yang akan diestimasi berukuran  $(p + 1) \times 1$ ,  $\tilde{\varepsilon}$  adalah vektor *error* random berukuran  $n \times 1$  berdistribusi normal, independen dengan mean nol dan variansi  $\sigma^2$ . Secara lengkap matriks dan vektor-vektor tersebut diberikan oleh :

$$\tilde{\mathbf{y}} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \mathbf{X} = \begin{pmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ 1 & x_{31} & x_{32} & \cdots & x_{3p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix}, \tilde{\boldsymbol{\beta}} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{pmatrix} \text{ dan } \tilde{\boldsymbol{\varepsilon}} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

Model regresi pada persamaan (2.7) adalah model regresi parametrik linear. Bentuk pola kurva regresi yang diketahui inilah merupakan ciri dari regresi parametrik dimana terdapat asumsi yang sangat kuat yaitu pola data mengikuti pola linear, kuadratik, polinomial derajat  $p$ , eksponen, dan pola-pola lainnya (Budiantara, 2009). Estimator parameter-parameter model diperoleh berdasarkan berbagai metode yang telah dikenal dalam Statistika, yaitu *Least Square* atau *Maximum Likelihood Estimator*, jika digunakan metode *least square* didapat estimator parameter :

$$\hat{\tilde{\boldsymbol{\beta}}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \tilde{\mathbf{y}} \quad (2.8)$$

## 2.4 Estimator Spline

*Spline* merupakan salah satu teknik estimasi regresi nonparametrik yang pertama kali dikembangkan oleh Whittaker pada tahun 1923 (Hardle, 1990). Regresi nonparametrik merupakan regresi yang sangat fleksibel dalam memodelkan pola data (Eubank, 1988). *Spline* juga memiliki kemampuan yang sangat baik untuk menangani data yang perilakunya berubah-ubah pada sub-sub interval tertentu disebabkan *Spline* merupakan model polinomial yang tersegmen/terpotong (Budiantara, dkk, 2009).

Diberikan data berpasangan  $(t_i, y_i)$ ,  $i = 1, 2, \dots, n$ , dimana pola hubungannya dapat dinyatakan dalam model regresi  $y_i = g(t_i) + \varepsilon_i$ . Kurva regresi  $g(t_i)$  dihampiri dengan kurva regresi *Spline truncated*. Secara umum, fungsi *Spline* derajat  $q$  adalah sembarang fungsi yang dapat disajikan dalam bentuk (Eubank, 1988) :

$$\begin{aligned} g(t_i) &= \theta_0 + \theta_1 t_i^1 + \dots + \theta_q t_i^q + \lambda_1 (t_i - K_1)_+^q + \lambda_2 (t_i - K_2)_+^q + \dots + \lambda_m (t_i - K_m)_+^q \\ &= \sum_{j=0}^q \theta_j t_i^j + \sum_{k=1}^m \lambda_k (t_i - K_k)_+^q \end{aligned} \quad (2.9)$$

dengan  $(t_i - K_k)_+^q = \begin{cases} (t_i - K_k)^q; t_i \geq K_k \\ 0; t_i < K_k \end{cases}$

$\theta_j$  adalah konstanta real,  $K_k$  adalah knot ke- $k$ ,  $k = 1, 2, \dots, m$ . Sehingga model regresinya menjadi :

$$y_i = g(t_i) + \varepsilon_i \quad (2.10)$$

$$y_i = \sum_{j=0}^q \theta_j t_i^j + \sum_{k=1}^m \lambda_k (t_i - K_k)_+^q + \varepsilon_i \quad (2.11)$$

dalam bentuk matriks dapat ditulis :

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} \theta_0 + \theta_1 t_1^1 + \dots + \theta_q t_1^q + \lambda_1 (t_1 - K_1)_+^q + \lambda_2 (t_1 - K_2)_+^q + \dots + \lambda_m (t_1 - K_m)_+^q \\ \theta_0 + \theta_1 t_2^1 + \dots + \theta_q t_2^q + \lambda_1 (t_2 - K_1)_+^q + \lambda_2 (t_2 - K_2)_+^q + \dots + \lambda_m (t_2 - K_m)_+^q \\ \vdots \\ \theta_0 + \theta_1 t_n^1 + \dots + \theta_q t_n^q + \lambda_1 (t_n - K_1)_+^q + \lambda_2 (t_n - K_2)_+^q + \dots + \lambda_m (t_n - K_m)_+^q \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

$$= \begin{bmatrix} 1 & t_1^1 & \dots & t_1^q & (t_1 - K_1)_+^q & (t_1 - K_2)_+^q & \dots & (t_1 - K_m)_+^q \\ 1 & t_2^1 & \dots & t_2^q & (t_2 - K_1)_+^q & (t_2 - K_2)_+^q & \dots & (t_2 - K_m)_+^q \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & t_n^1 & \dots & t_n^q & (t_n - K_1)_+^q & (t_n - K_2)_+^q & \dots & (t_n - K_m)_+^q \end{bmatrix} \begin{bmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_q \\ \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_m \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

atau dapat ditulis dalam notasi matriks :

$$\tilde{y} = \mathbf{G}(k) \tilde{\theta} + \tilde{\varepsilon} \quad (2.12)$$

dengan :

$$\tilde{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad \mathbf{G}(k) = \begin{bmatrix} 1 & t_1^1 & \dots & t_1^q & (t_1 - K_1)_+^q & (t_1 - K_2)_+^q & \dots & (t_1 - K_m)_+^q \\ 1 & t_2^1 & \dots & t_2^q & (t_2 - K_1)_+^q & (t_2 - K_2)_+^q & \dots & (t_2 - K_m)_+^q \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & t_n^1 & \dots & t_n^q & (t_n - K_1)_+^q & (t_n - K_2)_+^q & \dots & (t_n - K_m)_+^q \end{bmatrix},$$

$$\tilde{\theta} = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_q \\ \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_m \end{bmatrix}, \text{ dan } \tilde{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

*Spline* merupakan salah satu model yang mempunyai interpretasi statistik dan visual yang khusus dan baik (Eubank, 1988). Dalam fungsi *Spline* terdapat titik knot yang merupakan titik perpaduan yang menunjukkan perubahan perilaku kurva pada selang yang berbeda (Hardle, 1990). Titik knot merupakan bagian yang sangat penting dalam regresi *Spline*. Oleh karena itu agar diperoleh *Spline* yang optimal perlu dipilih titik knot yang terbaik, berapa jumlahnya dan dimana letak titik-titik knot tersebut. Terdapat 2 strategi untuk menyelesaikan permasalahan ini, strategi pertama adalah memilih banyaknya knot yang relatif sedikit, sedangkan strategi kedua adalah kebalikannya yakni menggunakan knot yang relatif banyak (Wand, 2000). Dari kedua hal tersebut yang paling baik adalah lebih mengarah pada alasan kesederhanaan model (*parsimony*). Salah satu metode untuk memilih titik knot optimal adalah dengan metode *Generalized Cross Validation* (Budiantara, 2000).

Estimator parameter-parameter model diperoleh berdasarkan berbagai metode yang telah dikenal dalam Statistika, yaitu *Least Square* atau *Maximum Likelihood Estimator*, jika digunakan metode *least square* didapat estimator parameter :

$$\hat{\theta} = [\mathbf{G}(k)^T \mathbf{G}(k)]^{-1} \mathbf{G}(k)^T \tilde{y} \quad (2.13)$$

dengan demikian estimasi dari kurva regresi  $\hat{g}(t_i)$  diperoleh :

$$\begin{aligned} \hat{g}(t_i) &= \mathbf{G}(k) \tilde{\theta} \\ &= \mathbf{G}(k) [\mathbf{G}(k)^T \mathbf{G}(k)]^{-1} \mathbf{G}(k)^T \tilde{y} \\ &= \mathbf{A}(k) \tilde{y} \end{aligned} \quad (2.14)$$

dengan :

$$\mathbf{A}(k) = \mathbf{G}(k) \left[ \mathbf{G}(k)^T \mathbf{G}(k) \right]^{-1} \mathbf{G}(k)^T$$

## 2.5 Estimator Kernel

Estimator kernel diperkenalkan pertama kali oleh Rosenblatt (1956) dan Parzen (1962) sehingga namanya disebut estimator densitas kernel Rosenblatt-Parzen (Hardle, 1990). Estimator kernel merupakan perkembangan dari estimator histogram. Estimator kernel merupakan estimator linier yang serupa dengan estimator lainnya, perbedaannya adalah metode kernel lebih khusus dalam penggunaan *bandwidth* (Eubank, 1988). Beberapa kelebihan estimator kernel adalah fleksibel, bentuk matematisnya mudah, dan dapat mencapai tingkat kekonvergenan yang relatif cepat (Budiantara dan Mulianah, 2007).

Diberikan data berpasangan  $(z_i, y_i)$ ,  $i = 1, 2, \dots, n$ , dimana pola hubungannya dapat dinyatakan dalam model regresi  $y_i = h(z_i) + \varepsilon_i$ . Kurva regresi  $h(z_i)$  dihampiri dengan kurva regresi kernel. Untuk mengestimasi fungsi regresi  $h(z_i)$  pada model regresi nonparametrik, Nadaraya dan Watson pada tahun 1964 mendefinisikan estimator regresi kernel sehingga disebut estimator Nadaraya-Watson (Hardle, 1994).

$$\hat{h}_\alpha(z) = n^{-1} \sum_{j=1}^n W_{\alpha j}(z) y_j \quad (2.15)$$

$$\text{dengan } W_{\alpha j}(z) = \frac{\frac{1}{\alpha} K\left(\frac{z - z_j}{\alpha}\right)}{n^{-1} \sum_{j=1}^n \frac{1}{\alpha} K\left(\frac{z - z_j}{\alpha}\right)}.$$

Dimana  $\hat{h}_\alpha(z)$  adalah fungsi taksiran regresi kernel,  $z$  adalah variabel prediktor yang nilainya tidak teramati namun akan digunakan untuk menaksir,  $z_i$  adalah variabel prediktor pada data ke- $i$ ,  $Y_i$  adalah variabel respon data ke- $i$ ,  $K$  merupakan fungsi kernel,  $n$  adalah banyak pengamatan dan  $\alpha$  adalah lebar *bandwidth*. Estimator kernel tergantung pada fungsi kernel  $K$  dan parameter *bandwidth*  $\alpha$ . Dalam bentuk matriks dapat ditulis :

$$\begin{aligned}
\begin{bmatrix} \hat{h}_\alpha(z_1) \\ \hat{h}_\alpha(z_2) \\ \vdots \\ \hat{h}_\alpha(z_n) \end{bmatrix} &= \begin{bmatrix} n^{-1}W_\alpha(z_1)y_1 + n^{-1}W_\alpha(z_1)y_2 + \dots + n^{-1}W_\alpha(z_1)y_n \\ n^{-1}W_\alpha(z_2)y_1 + n^{-1}W_\alpha(z_2)y_2 + \dots + n^{-1}W_\alpha(z_2)y_n \\ \vdots \\ n^{-1}W_\alpha(z_n)y_1 + n^{-1}W_\alpha(z_n)y_2 + \dots + n^{-1}W_\alpha(z_n)y_n \end{bmatrix} \\
&= \begin{bmatrix} n^{-1}W_\alpha(z_1) & n^{-1}W_\alpha(z_1) & \dots & n^{-1}W_\alpha(z_1) \\ n^{-1}W_\alpha(z_2) & n^{-1}W_\alpha(z_2) & \dots & n^{-1}W_\alpha(z_2) \\ \vdots & \vdots & \ddots & \vdots \\ n^{-1}W_\alpha(z_n) & n^{-1}W_\alpha(z_n) & \dots & n^{-1}W_\alpha(z_n) \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}
\end{aligned}$$

atau dapat ditulis dalam notasi matriks :

$$\hat{\tilde{h}}_\alpha(z) = \mathbf{D}(\alpha) \tilde{y} \quad (2.16)$$

Ada beberapa jenis fungsi kernel, diantaranya sebagai berikut (Hardle, 1990) :

a) Kernel Uniform

$$K(x) = \frac{1}{2}, \quad |x| \leq 1$$

b) Kernel Triangle

$$K(x) = 1 - |x|, \quad |x| \leq 1$$

c) Kernel Epanechnikov

$$K(x) = \frac{3}{4}(1 - x^2), \quad |x| \leq 1$$

d) Kernel Kuartik

$$K(x) = \frac{15}{16}(1 - x^2)^2, \quad |x| \leq 1$$

e) Kernel Triweight

$$K(x) = \frac{35}{32}(1 - x^2)^3, \quad |x| \leq 1$$

f) Kernel Cosinus

$$K(x) = \frac{\pi}{4} \cos\left(\frac{\pi}{2}x\right), \quad |x| \leq 1$$

g) Kernel Gaussian

$$K(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \quad -\infty < x < \infty$$

Adapun fungsi Kernel yang digunakan adalah fungsi Kernel Gaussian.

## 2.6 Generalized Cross Validation (GCV)

Perolehan *Spline* optimal dan kemulusan kurvanya bergantung pada pemilihan titik–titik knot  $k$ . Titik knot  $k$  merupakan perpaduan bersama antara perubahan fungsi pada interval yang berlainan. Sementara itu pada kernel bergantung pada *bandwidth*. *Bandwidth*  $\alpha$  adalah parameter pemulus (*smoothing*) yang berfungsi untuk mengontrol kemulusan dari kurva yang diestimasi. Titik knot maupun *bandwidth* yang terlalu kecil akan menghasilkan kurva yang *under-smoothing* yaitu sangat kasar dan sangat fluktuatif, dan sebaliknya titik knot maupun *bandwidth* yang terlalu besar/lebar akan menghasilkan kurva yang *over-smoothing* yaitu sangat mulus, tetapi tidak sesuai dengan pola data (Hardle, 1994). Oleh karena itu perlu dipilih titik knot maupun *bandwidth* yang optimal. Salah satu metode untuk mendapatkan  $k$  dan  $\alpha$  optimal adalah dengan menggunakan kriteria *Generalized Cross Validation* (GCV). Pemilihan titik knot  $k$  optimal pada komponen *Spline truncated* menggunakan GCV didefinisikan sebagai berikut :

$$GCV(k) = \frac{MSE(k)}{\left(n^{-1}tr(\mathbf{I} - \mathbf{A}(k))\right)^2} \quad (2.17)$$

dimana :

$$MSE(k) = n^{-1} \sum_{i=1}^n (y_i - \hat{y}_i)^2, \text{ dan matriks } \mathbf{A}(k) \text{ seperti pada persamaan (2.14).}$$

Sementara itu, Pemilihan *bandwidth*  $\alpha$  optimal pada komponen Kernel menggunakan GCV didefinisikan sebagai berikut :

$$GCV(\alpha) = \frac{MSE(\alpha)}{\left(n^{-1}tr(\mathbf{I} - \mathbf{D}(\alpha))\right)^2} \quad (2.18)$$

dimana :

$$MSE(\alpha) = n^{-1} \sum_{i=1}^n (y_i - \hat{y}_i)^2, \text{ dan matriks } \mathbf{D}(\alpha) \text{ seperti pada persamaan (2.16).}$$

## 2.7 Koefisien Determinasi $R^2$

Salah satu tujuan analisis regresi adalah mendapatkan model terbaik yang mampu menjelaskan hubungan antara variabel prediktor dan variabel respon berdasarkan kriteria tertentu. Salah satu kriteria yang digunakan dalam pemilihan

model terbaik adalah dengan menggunakan koefisien determinasi  $R^2$ . Secara umum semakin besar nilai  $R^2$ , maka semakin baik pula model yang didapatkan. Menurut Eubank (1988), koefisien determinasi didefinisikan sebagai berikut :

$$R^2 = \frac{(\hat{y} - \bar{y})^T (\hat{y} - \bar{y})}{(y - \bar{y})^T (y - \bar{y})} \quad (2.19)$$

## 2.8 Produksi Padi

Untuk mencapai kemandirian dan kedaulatan pangan, ketersediaan komoditas tanaman pangan untuk konsumsi dalam negeri harus terjamin. Komoditas yang nilainya paling strategis adalah padi. Padi merupakan tanaman penghasil beras yang merupakan makanan pokok sebagian besar penduduk Indonesia (BPS, 2014). Karena demikian strategisnya padi termasuk dalam salah satu sasaran utama pemerintah dalam target swasembada sebagaimana tertuang dalam RPJM 2015-2019. Sebagai komoditas terpenting di dalam pembangunan pertanian maka berbagai usaha dilakukan untuk meningkatkan produksi padi.

Pertanian mempunyai kontribusi penting terhadap perekonomian yaitu kontribusi produk dalam sumbangannya terhadap Produk Domestik Regional Bruto (PDRB) dan kontribusi pasar dibandingkan dengan usaha-usaha lainnya. Menurut publikasi Pusdatin Kementerian Pertanian Republik Indonesia, sekitar 70 persen dari total produk domestik bruto Indonesia yang disumbangkan oleh sektor pertanian berasal dari pertanian padi. Pada tahun 2010, pertanian padi berkontribusi terhadap tujuh persen dari total produk domestik bruto Indonesia secara keseluruhan. Peran penting lainnya adalah dalam penyediaan kebutuhan pangan penduduk, apalagi dengan semakin meningkatnya jumlah penduduk yang berarti bahwa kebutuhan akan pangan juga semakin meningkat. Oleh karena itu peningkatan produksi padi perlu dilakukan dan perlu diperhatikan secara intensif pula apa saja yang mempengaruhi faktor produksinya.

Indonesia yang pernah swasembada beras tahun 1985, 1986, dan 1993 serta merupakan penghasil padi nomor tiga terbesar di dunia, ternyata saat ini Indonesia malah menjadi negara pengimpor beras terbesar di dunia (14% dari yang diperdagangkan dunia). Tingginya ketergantungan Indonesia terhadap impor beras dunia merupakan salah satu alasan juga mengapa upaya peningkatan



produksi beras nasional perlu dilakukan. Produksi padi nasional mengalami kenaikan hanya 2 % dari tahun 2012 ke tahun 2013, Namun pada 2014 mengalami penurunan menjadi 69,87 ribu ton gabah kering giling (GKG).

Pulau Jawa merupakan sentra produksi padi terbesar di Indonesia sehingga tidak mengherankan produksi padi tertinggi terdapat di Pulau Jawa. Dibandingkan dengan produksi padi di luar pulau jawa, pulau jawa selalu menghasilkan produksi yang lebih tinggi. Berdasarkan data BPS, provinsi penghasil produksi padi terbesar adalah provinsi Jawa Barat dan Jawa Timur. Sementara itu Jawa Tengah menghasilkan produksi padi yang lebih sedikit dibanding kedua provinsi tersebut. Bahkan dalam kurun antara tahun 2010-2014, provinsi Jawa Tengah mengalami dua kali penurunan hasil produksi padi, yakni pada tahun 2011 dan 2014. Selain itu, produksi padi Jawa Tengah masih jauh berada dibawah produksi Jawa Barat dan Jawa Timur. Oleh karena itu diperlukan analisis untuk mengetahui faktor yang berpengaruh terhadap produksi padi di Jawa Tengah. Produksi padi sangat ditentukan oleh penggunaan faktor-faktor produksi seperti pupuk, tenaga kerja, benih, dan pestisida.

Salah satu faktor yang mempengaruhi produksi padi adalah luas lahan, Pada penelitian yang dilakukan oleh Susianto (2008) diketahui bahwa luas lahan berhubungan positif terhadap produksi padi, dimana ketika luas lahan meningkat maka produksi padi akan meningkat. Pada penelitian lain yang dilakukan oleh Asriani (2013) juga menyimpulkan hal yang sama. Sehingga pada penelitian ini padi diasumsikan berhubungan linear. Hubungan tersebut selanjutnya akan didukung oleh *scatter plot* dari data yang akan digunakan. Sementara itu variabel lain yang mempengaruhi produksi padi diketahui berhubungan nonparametrik. Oleh karena itu, pada penelitian ini menggunakan regresi semiparametrik.

Selain itu, menurut Triyanto (2006) yang berpengaruh terhadap produksi padi adalah pupuk. Pupuk menjadi penting karena merupakan salah satu *input* dari proses produksi padi. Prinsip pemupukan yang benar adalah menambahkan ke dalam tanah unsur-unsur yang kurang, tetapi dibutuhkan oleh tanaman. Penggunaan pupuk dalam jumlah sedikit maupun dalam jumlah banyak akan memberikan pengaruh yang sama terhadap kesuburan tanah yang selanjutnya berakibat pada produksi padi. Hal ini juga sejalan dengan beberapa penelitian

sebelumnya yang menemukan adanya hubungan antara penggunaan pupuk terhadap produksi padi (Zulkarnain, 2004) dan (Zuriani, 2012).

Sama halnya dengan pupuk, penggunaan pestisida dalam jumlah banyak juga tidak baik. Menurut Subiyakto (1992) penggunaan pestisida yang banyak dan dilakukan terus menerus akan menyebabkan kerugian, yaitu pencemaran pada lingkungan pertanian padi yang berakibat pada hasil produksi. Hal ini disebabkan oleh pestisida yang merupakan obat-obatan atau senyawa kimia yang umumnya bersifat racun. Ini berarti penggunaan pestisida dalam jumlah sedikit maupun dalam jumlah banyak akan memberikan pengaruh yang sama terhadap produksi padi.

Penggunaan benih yang baik merupakan salah satu cara peningkatan produksi, namun pemenuhan kebutuhan benih bermutu di Jawa Tengah baru mencapai 80% (Dispertan, 2004) sehingga sisanya dipenuhi dari penggunaan benih yang kurang bermutu. Namun karena keterbatasan data yang dimiliki, maka pada penelitian ini, bukan terkait mutu benih yang digunakan melainkan terkait dengan banyaknya benih yang digunakan. Beberapa penelitian sebelumnya menemukan adanya pengaruh penggunaan benih terhadap produksi padi (Triyanto, 2006).

Faktor lain yang diduga berpengaruh adalah tenaga kerja. Tenaga kerja merupakan salah satu faktor produksi yang utama dalam usaha tani. Tenaga kerja dapat diperoleh dari dalam maupun dari luar rumah tangga. Seperti halnya pupuk, pestisida dan benih, penggunaan tenaga kerja juga harus optimal. Menurut Asriani (2013) penggunaan tenaga kerja yang terlalu sedikit ataupun terlalu banyak cenderung tidak efektif dan tidak efisien. Hal ini sejalan dengan penelitian sebelumnya yang menemukan hubungan antara penggunaan tenaga kerja berpengaruh terhadap produksi padi (Zulkarnain, 2004) dan (Zuriani, 2013).

Pada penelitian sebelumnya untuk variabel penggunaan pupuk, penggunaan pestisida, penggunaan benih dan penggunaan tenaga kerja diketahui berhubungan nonparametrik dengan produksi padi. Susianto (2008) mendekati dengan regresi nonparametrik kernel dengan nilai  $R^2$  yang diperoleh adalah sebesar 82,47. Sementara itu, Asriani (2013) mendekati dengan regresi nonparametrik deret fourier. Namun dengan dasar bahwa pola data pada keempat

variabel tersebut berbeda maka dapat dipertimbangkan untuk menggunakan estimator yang berbeda pula yang sesuai dengan pola data. Sehingga dengan identifikasi awal menggunakan *scatter plot* dari masing-masing variabel prediktor terhadap variabel respon, maka pada penelitian ini dapat digunakan estimator campuran *Spline truncated* dan kernel. *Spline* digunakan untuk data dengan pola hubungan yang berubah-ubah pada sub-sub interval tertentu. Sementara itu, Kernel digunakan untuk data dengan tanpa ada pola tertentu.

*Halaman ini sengaja dikosongkan*



## BAB III

### METODOLOGI PENELITIAN

#### 3.1 Sumber Data

Data yang akan digunakan dalam penelitian ini adalah data sekunder yang diperoleh dari publikasi BPS. Data yang diambil merupakan data yang berkaitan dengan produksi padi di Jawa Tengah tahun 2008, dengan unit penelitian adalah rumah tangga yang sumber pendapatan utamanya berasal dari tanaman padi (sebanyak 691 rumah tangga usaha tani padi).

#### 3.2 Variabel Penelitian dan Struktur Data

Variabel-variabel yang akan digunakan dalam penelitian ini yaitu pada Tabel 3.1.

Tabel 3.1 Variabel Penelitian

Variabel		Nama Variabel	Satuan
Respon	$Y_1$	Produksi Padi	Ton
Prediktor	$X_1$	Luas Panen	Ha
	$X_2$	Penggunaan Pupuk	Kg / Ha
	$X_3$	Penggunaan Pestisida	liter / Ha
	$X_4$	Penggunaan Benih	Kg / Ha
	$X_5$	Penggunaan Tenaga Kerja	Orang/Ha

Adapun struktur data pada penelitian ini seperti terlihat pada Tabel 3.2.

Tabel 3.2 Struktur Data Penelitian

Data	$X_1$	$X_2$	$X_3$	$X_4$	$X_5$	$Y$
1	$X_{1(1)}$	$X_{2(1)}$	$X_{3(1)}$	$X_{4(1)}$	$X_{5(1)}$	$Y_{1(1)}$
2	$X_{1(2)}$	$X_{2(2)}$	$X_{3(2)}$	$X_{4(2)}$	$X_{5(2)}$	$Y_{1(2)}$
3	$X_{1(3)}$	$X_{2(3)}$	$X_{3(3)}$	$X_{4(3)}$	$X_{5(3)}$	$Y_{1(3)}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
691	$X_{1(691)}$	$X_{2(691)}$	$X_{3(691)}$	$X_{4(691)}$	$X_{5(691)}$	$Y_{1(691)}$

### 3.3 Definisi Operasional

Adapun definisi operasional dari variabel-variabel yang digunakan dalam penelitian ini adalah sebagai berikut :

a. Produksi padi (Ton)

Produksi padi adalah produksi kotor dalam bentuk standar yang berbentuk gabah kering giling (GKG) yang dihasilkan oleh masing-masing rumah tangga petani dalam satuan ton.

b. Luas panen (Ha)

Luas panen adalah luas tanaman padi sawah yang dipungut hasilnya (paling sedikit sampai dengan 10%) oleh masing-masing rumah tangga petani dalam satuan hektar (Ha).

c. Penggunaan pupuk (kg/Ha)

Penggunaan pupuk adalah jumlah pupuk yang digunakan oleh masing-masing rumah tangga petani dalam satuan kilogram per hektar(kg/Ha).

d. Penggunaan pestisida (liter/Ha)

Penggunaan pestisida adalah jumlah pestisida yang digunakan oleh masing-masing rumah tangga petani dalam satuan liter perhektar (liter/ha).

e. Penggunaan benih (kg/Ha)

Penggunaan benih adalah jumlah benih yang digunakan oleh masing-masing rumah tangga petani dalam satuan kilogram per hektar (kg/Ha).

f. Penggunaan tenaga kerja (orang/Ha)

Penggunaan tenaga kerja adalah jumlah tenaga kerja yang digunakan oleh masing-masing rumah tangga tani dalam satuan orang per hektar (org/Ha).

### 3.4 Metode Penelitian

Secara umum langkah-langkah yang akan dilakukan dalam penelitian ini adalah :

1. Mengestimasi bentuk kurva regresi semiparametrik campuran *Spline truncated* dan kernel, dengan langkah-langkah sebagai berikut :

a. Diberikan respon  $y_i$  dengan variabel prediktor komponen parametrik  $x_i$ .

Sedangkan variabel prediktor  $t_i$  dan  $z_i$  merupakan komponen nonparametrik. Hubungan antara  $x_i$ ,  $t_i$  dan  $z_i$  mengikuti model regresi :

$$y_i = \mu(x_i, t_i, z_i) + \varepsilon_i, \quad i = 1, 2, \dots, n$$

- b. Mendekomposisikan kurva regresi  $\mu$  menjadi bentuk additif :

$$\mu(x_i, t_i, z_i) = f(x_i) + g(t_i) + h(z_i)$$

- c. Menghampiri fungsi  $f(x_i)$  menggunakan fungsi linear,  $g(t_i)$  menggunakan fungsi *Spline truncated linear*, dan  $h(z_i)$  menggunakan fungsi kernel.
- d. Mencari bentuk estimasi kurva regresi  $\mu(x_i, t_i, z_i)$  menggunakan metode OLS melalui optimasi :

$$Q(f, g, h) = \underset{f, g, h}{\text{Min}} \left\{ \left\| \tilde{y} - \tilde{f}(x) - \tilde{g}(t) - \tilde{h}(z) \right\|^2 \right\}$$

- e. Menyatakan kurva regresi  $\tilde{f}$ ,  $\tilde{g}$  dan  $\tilde{h}$  dalam bentuk :

$$\tilde{f}(x) = \mathbf{X}\tilde{\beta}, \quad \tilde{g}(t) = \mathbf{G}(k)\tilde{\theta} \quad \text{dan} \quad \tilde{h}(z) = \mathbf{D}(\alpha)\tilde{y}$$

$$\text{dengan } \mathbf{X} = \begin{pmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ 1 & x_{31} & x_{32} & \cdots & x_{3p} \\ & & & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix}, \quad \tilde{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{pmatrix},$$

$$\mathbf{G}(k) = \begin{bmatrix} t_1 & (t_1 - K_1)_+ & \cdots & (t_1 - K_m)_+ \\ t_2 & (t_2 - K_1)_+ & \cdots & (t_2 - K_m)_+ \\ \vdots & \vdots & \ddots & \vdots \\ t_n & (t_n - K_1)_+ & \cdots & (t_n - K_m)_+ \end{bmatrix}, \quad \tilde{\theta} = \begin{bmatrix} \theta_1 \\ \lambda_1 \\ \vdots \\ \lambda_m \end{bmatrix}$$

$$\mathbf{D}(\alpha) = \begin{bmatrix} n^{-1}W_\alpha(z_1) & n^{-1}W_\alpha(z_1) & \cdots & n^{-1}W_\alpha(z_1) \\ n^{-1}W_\alpha(z_2) & n^{-1}W_\alpha(z_2) & \cdots & n^{-1}W_\alpha(z_2) \\ \vdots & \vdots & \ddots & \vdots \\ n^{-1}W_\alpha(z_n) & n^{-1}W_\alpha(z_n) & \cdots & n^{-1}W_\alpha(z_n) \end{bmatrix}, \quad \tilde{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$



f. Menyelesaikan optimasi :

$$Q(\tilde{\beta}, \tilde{\theta}) = \underset{\tilde{\beta}, \tilde{\theta}}{\text{Min}} \left\| \tilde{y} - \mathbf{X}\tilde{\beta} - \mathbf{G}(k)\tilde{\theta} - \mathbf{D}(\alpha)\tilde{y} \right\|^2$$

dengan menggunakan derivatif parsial.

g. Menyamakan derivatif parsial tersebut dengan nol :

$$\frac{\partial Q(\tilde{\beta}, \tilde{\theta})}{\partial \tilde{\beta}} = 0$$

$$\frac{\partial Q(\tilde{\beta}, \tilde{\theta})}{\partial \tilde{\theta}} = 0$$

h. Mendapatkan estimasi campuran

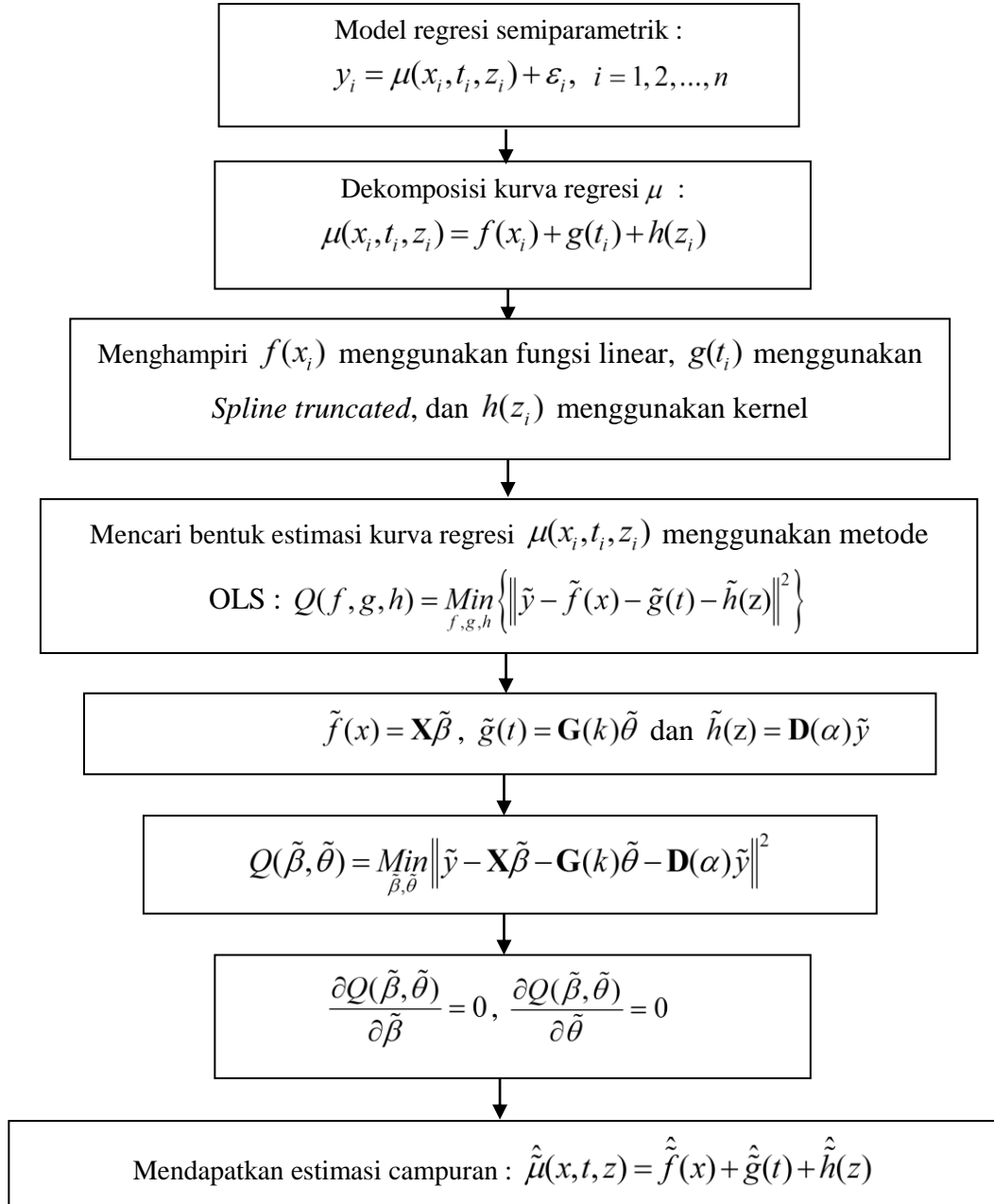
$$\hat{\mu}(x, t, z) = \hat{f}(x) + \hat{g}(t) + \hat{h}(z)$$

2. Mengaplikasikan estimasi campuran yang diperoleh untuk memodelkan produksi padi di Jawa Tengah, dengan langkah-langkah sebagai berikut :

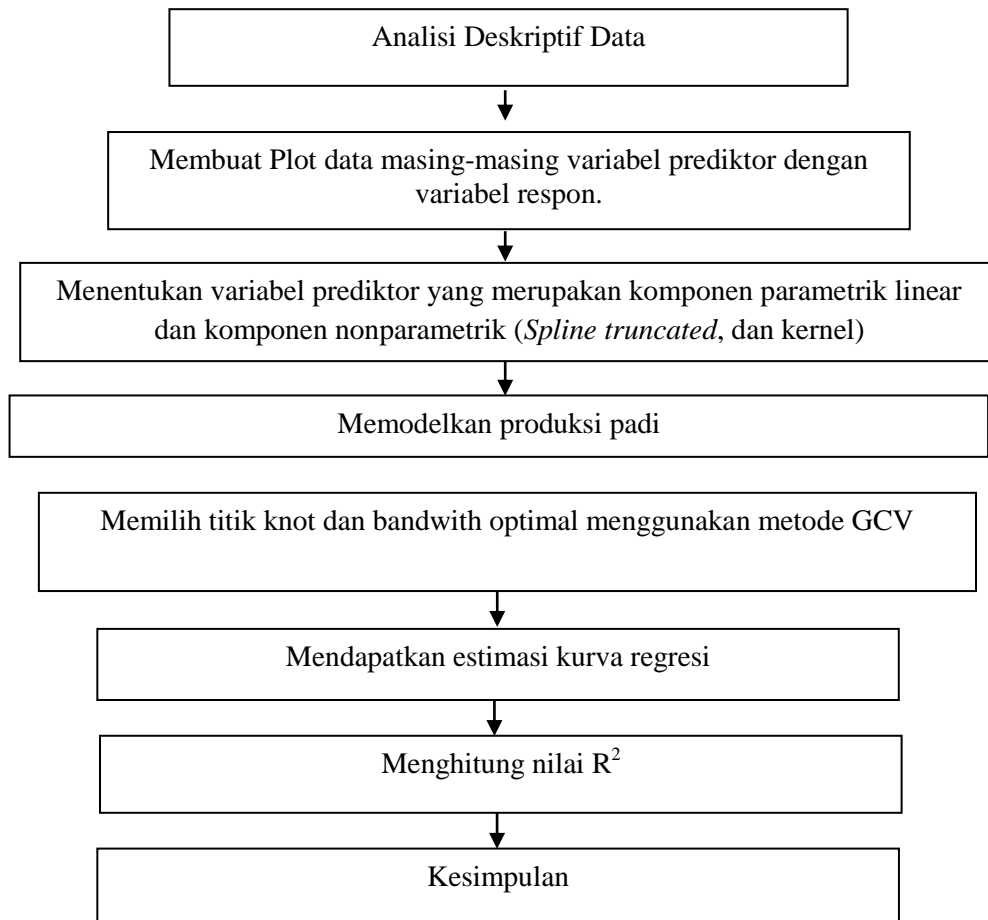
- a. Melakukan analisis deskriptif data untuk mengetahui gambaran umum tentang data produksi padi di Jawa Tengah, dengan variabel-variabel prediktornya.
- b. Membuat plot data antara variabel respon dengan masing-masing variabel prediktor.
- c. Menentukan variabel prediktor yang merupakan komponen parametrik, komponen nonparametrik yang didekati dengan fungsi *Spline truncated* dan yang didekati dengan fungsi kernel.
- d. Memodelkan produksi padi di Jawa Tengah dengan menggunakan pendekatan semiparametrik campuran *Spline truncated* dan kernel.
- e. Memilih titik knot dan *bandwidth* optimal menggunakan metode GCV.
- f. Mendapatkan estimasi kurva regresi semiparametrik campuran *Spline truncated* dan kernel.
- g. Menghitung  $R^2$

h. Menentukan kesimpulan yang sesuai dengan tujuan penelitian.

Langkah-langkah analisis dalam penelitian ini dapat digambarkan pada diagram alir seperti Gambar 3.1 dan Gambar 3.2 :



Gambar 3.1 Diagram Alir Langkah Analisis untuk Tujuan Pertama



Gambar 3.2 Diagram Alir Langkah Analisis untuk Tujuan Kedua

## BAB IV

### HASIL DAN PEMBAHASAN

Berdasarkan tujuan penelitian, pada bagian ini dibahas mengenai model dan estimasi regresi semiparametrik campuran *Spline truncated* linear dan kernel. Selanjutnya hasil estimasi akan diterapkan pada pemodelan data produksi padi di Jawa Tengah Tahun 2008.

#### 4.1 Model Regresi Semiparametrik Campuran

Berikut ini dibahas mengenai model regresi semiparametrik campuran *Spline truncated* linear dan kernel.

##### 4.1.1 Model Regresi Semiparametrik Campuran *Spline Truncated* Linear dan Kernel

Diberikan data berpasangan  $(x_i, t_i, z_i, y_i)$ ,  $i = 1, 2, \dots, n$ , yang diasumsikan mengikuti model regresi semiparametrik :

$$y_i = \mu(x_i, t_i, z_i) + \varepsilon_i \quad (4.1)$$

kurva regresi  $\mu$  pada persamaan (4.1) diasumsikan bersifat additif sehingga dapat ditulis menjadi :

$$\mu(x_i, t_i, z_i) = f(x_i) + g(t_i) + h(z_i) \quad (4.2)$$

Bentuk pola hubungan antara variabel respon  $y_i$  dengan variabel prediktor  $x_i$  diasumsikan mengikuti pola linear. Sementara itu, pola hubungan antara variabel respon  $y_i$  dengan variabel prediktor  $t_i$ ,  $z_i$  diasumsikan tidak diketahui dan hanya diasumsikan *smooth* dalam arti kontinu dan differensiabel, dimana bentuk pola hubungan variabel respon  $y_i$  dengan variabel prediktor  $t_i$  diasumsikan berubah-ubah pada sub-sub interval tertentu. Sementara itu, Bentuk pola hubungan variabel respon  $y_i$  dengan variabel prediktor  $z_i$  diasumsikan tidak memiliki pola tertentu. Sehingga secara teoritis kurva regresi  $g(t_i)$  didekati dengan fungsi *Spline truncated* linear. Sementara itu, kurva regresi  $h(z_i)$  dapat dihipotesis dengan fungsi kernel. Dengan demikian, kurva campuran  $\mu(x_i, t_i, z_i)$  secara keseluruhan bisa didekati dengan kurva regresi semiparametrik, dengan

$f(x_i)$  merupakan komponen parametrik. Sementara itu, komponen nonparametrik didekati dengan kurva campuran *Spline truncated* linear dan kernel, dengan  $g(t_i)$  merupakan komponen nonparametrik *Spline truncated* dan  $h(z_i)$  merupakan komponen nonparametrik kernel.

Komponen parametrik linear disajikan dalam bentuk :

$$f(x_i) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} \quad (4.3)$$

Dari persamaan fungsi linear sederhana di atas, berlaku untuk  $i = 1$  sampai  $n$ , diperoleh :

$$\begin{bmatrix} f(x_1) \\ f(x_2) \\ \vdots \\ f(x_n) \end{bmatrix} = \begin{bmatrix} \beta_0 + \beta_1 x_{11} + \dots + \beta_p x_{1p} \\ \beta_0 + \beta_1 x_{21} + \dots + \beta_p x_{2p} \\ \vdots \\ \beta_0 + \beta_1 x_{n1} + \dots + \beta_p x_{np} \end{bmatrix}$$

Selanjutnya dapat ditulis menjadi :

$$\begin{bmatrix} f(x_1) \\ f(x_2) \\ \vdots \\ f(x_n) \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & \dots & x_{1p} \\ 1 & x_{21} & \dots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \dots & x_{np} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}$$

Dalam notasi matriks dapat ditulis menjadi :

$$\tilde{f}(x) = \mathbf{X}\tilde{\beta}, \quad (4.4)$$

dimana :

$$\tilde{f}(x) = \begin{bmatrix} f(x_1) \\ f(x_2) \\ \vdots \\ f(x_n) \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} 1 & x_{11} & \dots & x_{1p} \\ 1 & x_{21} & \dots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \dots & x_{np} \end{bmatrix} \quad \text{dan} \quad \tilde{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}$$

Sementara itu, komponen kurva regresi nonparametrik *Spline truncated* linear dengan knot  $k_1, k_2, \dots, k_m$  dapat disajikan menjadi :

$$\begin{aligned} g(t_i) &= \theta_1 t_i + \sum_{k=1}^m \lambda_k (t_i - K_k)_+ \\ &= \theta_1 t_i + \lambda_1 (t_i - K_1)_+ + \dots + \lambda_m (t_i - K_m)_+ \end{aligned} \quad (4.5)$$

dengan :

$$(t_i - K_k)_+ = \begin{cases} (t_i - K_k)_+ & , t_i \geq K_k \\ 0 & , t_i < K_k \end{cases}$$

dimana  $\theta_1, \lambda_1, \dots, \lambda_m$  merupakan parameter-parameter yang tidak diketahui,  $K_1, K_2, \dots, K_m$  merupakan titik knot dimana  $K_1 < K_2 < \dots < K_m$ . Dari persamaan fungsi *Spline truncated* di atas, untuk  $i = 1$  sampai  $n$ , didapat :

$$\begin{bmatrix} g(t_1) \\ g(t_2) \\ \vdots \\ g(t_n) \end{bmatrix} = \begin{bmatrix} \theta_1 t_1 + \lambda_1 (t_1 - K_1)_+ + \dots + \lambda_m (t_1 - K_m)_+ \\ \theta_1 t_2 + \lambda_1 (t_2 - K_1)_+ + \dots + \lambda_m (t_2 - K_m)_+ \\ \vdots \\ \theta_1 t_n + \lambda_1 (t_n - K_1)_+ + \dots + \lambda_m (t_n - K_m)_+ \end{bmatrix}$$

Selanjutnya dapat ditulis menjadi :

$$\begin{bmatrix} g(t_1) \\ g(t_2) \\ \vdots \\ g(t_n) \end{bmatrix} = \begin{bmatrix} t_1 & (t_1 - K_1)_+ & \dots & (t_1 - K_m)_+ \\ t_2 & (t_2 - K_1)_+ & \dots & (t_2 - K_m)_+ \\ \vdots & \vdots & \ddots & \vdots \\ t_n & (t_n - K_1)_+ & \dots & (t_n - K_m)_+ \end{bmatrix} \begin{bmatrix} \theta_1 \\ \lambda_1 \\ \vdots \\ \lambda_m \end{bmatrix}$$

Dalam notasi matriks dapat ditulis menjadi :

$$\tilde{g}(t) = \mathbf{G}(k) \tilde{\theta}, \quad (4.6)$$

dimana :

$$\tilde{g}(t) = \begin{bmatrix} g(t_1) \\ g(t_2) \\ \vdots \\ g(t_n) \end{bmatrix}, \quad \mathbf{G}(k) = \begin{bmatrix} t_1 & (t_1 - K_1)_+ & \dots & (t_1 - K_m)_+ \\ t_2 & (t_2 - K_1)_+ & \dots & (t_2 - K_m)_+ \\ \vdots & \vdots & \ddots & \vdots \\ t_n & (t_n - K_1)_+ & \dots & (t_n - K_m)_+ \end{bmatrix}, \quad \tilde{\theta} = \begin{bmatrix} \theta_1 \\ \lambda_1 \\ \vdots \\ \lambda_m \end{bmatrix}$$

Sementara itu, komponen kurva regresi nonparametrik kernel didefinisikan oleh :

$$\hat{h}_\alpha(z_i) = n^{-1} \sum_{i=1}^n W_{\alpha i}(z_i) y_i, \quad (4.7)$$

dimana :

$$W_{\alpha i}(z_i) = \frac{\frac{1}{\alpha} K\left(\frac{z_i - z_j}{\alpha}\right)}{n^{-1} \sum_{j=1}^n \frac{1}{\alpha} K\left(\frac{z_i - z_j}{\alpha}\right)}$$

dimana  $\hat{h}_\alpha(z_i)$  merupakan estimator kurva regresi kernel Nadaraya-Watson, parameter  $\alpha$  merupakan parameter *bandwidth*, fungsi kernel  $K$  yang digunakan adalah kernel Gaussian. Dari persamaan fungsi kernel di atas, untuk  $i = 1$  sampai  $n$ , didapat persamaan :

$$\begin{bmatrix} \hat{h}_\alpha(z_1) \\ \hat{h}_\alpha(z_2) \\ \vdots \\ \hat{h}_\alpha(z_n) \end{bmatrix} = \begin{bmatrix} n^{-1} \sum_{i=1}^n W_{\alpha i}(z_1) y_i \\ n^{-1} \sum_{i=1}^n W_{\alpha i}(z_2) y_i \\ \vdots \\ n^{-1} \sum_{i=1}^n W_{\alpha i}(z_n) y_i \end{bmatrix} = \begin{bmatrix} n^{-1} W_{\alpha 1}(z_1) y_1 + \cdots + n^{-1} W_{\alpha n}(z_1) y_n \\ n^{-1} W_{\alpha 1}(z_2) y_1 + \cdots + n^{-1} W_{\alpha n}(z_2) y_n \\ \vdots \\ n^{-1} W_{\alpha 1}(z_n) y_1 + \cdots + n^{-1} W_{\alpha n}(z_n) y_n \end{bmatrix}$$

Selanjutnya dapat juga ditulis menjadi :

$$\begin{bmatrix} \hat{h}_\alpha(z_1) \\ \hat{h}_\alpha(z_2) \\ \vdots \\ \hat{h}_\alpha(z_n) \end{bmatrix} = \begin{bmatrix} n^{-1} W_{\alpha 1}(z_1) & n^{-1} W_{\alpha 2}(z_1) & \cdots & n^{-1} W_{\alpha n}(z_1) \\ n^{-1} W_{\alpha 1}(z_2) & n^{-1} W_{\alpha 2}(z_2) & \cdots & n^{-1} W_{\alpha n}(z_2) \\ \vdots & \vdots & \ddots & \vdots \\ n^{-1} W_{\alpha 1}(z_n) & n^{-1} W_{\alpha 2}(z_n) & \cdots & n^{-1} W_{\alpha n}(z_n) \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

dalam notasi matriks dapat ditulis menjadi :

$$\hat{\mathbf{h}}(z) = \mathbf{D}(\alpha) \tilde{\mathbf{y}}, \quad (4.8)$$

dimana :

$$\hat{\mathbf{h}}(z) = \begin{bmatrix} \hat{h}_\alpha(z_1) \\ \hat{h}_\alpha(z_2) \\ \vdots \\ \hat{h}_\alpha(z_n) \end{bmatrix}, \quad \mathbf{D}(\alpha) = \begin{bmatrix} n^{-1} W_{\alpha 1}(z_1) & n^{-1} W_{\alpha 2}(z_1) & \cdots & n^{-1} W_{\alpha n}(z_1) \\ n^{-1} W_{\alpha 1}(z_2) & n^{-1} W_{\alpha 2}(z_2) & \cdots & n^{-1} W_{\alpha n}(z_2) \\ \vdots & \vdots & \ddots & \vdots \\ n^{-1} W_{\alpha 1}(z_n) & n^{-1} W_{\alpha 2}(z_n) & \cdots & n^{-1} W_{\alpha n}(z_n) \end{bmatrix}, \quad \tilde{\mathbf{y}} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

Berdasarkan uraian untuk komponen parametrik dan nonparametrik di atas, maka model regresi semiparametrik campuran *Spline truncated* dan kernel dapat dinyatakan menjadi :

$$\begin{aligned} \tilde{\mathbf{y}} &= f(x) + g(t) + \hat{\mathbf{h}}(z) + \tilde{\varepsilon} \\ &= X \tilde{\beta} + G(k) \tilde{\theta} + D(\alpha) \tilde{\mathbf{y}} + \tilde{\varepsilon} \end{aligned} \quad (4.9)$$

#### 4.1.2 Model Regresi Semiparametrik Campuran *Spline Truncated* dan Kernel Multivariabel

Diberikan data berpasangan  $(x_i, t_{1i}, t_{2i}, \dots, t_{pi}, z_{1i}, z_{2i}, \dots, z_{qi}, y_i)$  yang diasumsikan mengikuti model regresi semiparametrik :

$$y_i = \mu(x_i, t_{1i}, t_{2i}, \dots, t_{pi}, z_{1i}, z_{2i}, \dots, z_{qi}) + \varepsilon_i \quad (4.10)$$

kurva regresi  $\mu$  pada persamaan (4.10) diasumsikan bersifat additif sehingga dapat ditulis menjadi :

$$\begin{aligned} \mu(x_i, t_{1i}, t_{2i}, \dots, t_{pi}, z_{1i}, z_{2i}, \dots, z_{qi}) = & f(x_i) + g_1(t_{1i}) + g_2(t_{2i}) + \dots + g_p(t_{pi}) + \\ & h_1(z_{1i}) + h_2(z_{2i}) + \dots + h_q(z_{qi}). \end{aligned}$$

Fungsi  $f(x_i)$  dihampiri dengan parametrik linear,  $g(t_i)$  dihampiri dengan fungsi *Spline truncated* linear dan  $h(z_i)$  dihampiri dengan fungsi kernel. Dengan demikian, kurva campuran  $\mu(x_i, t_{1i}, t_{2i}, \dots, t_{pi}, z_{1i}, z_{2i}, \dots, z_{qi})$  secara keseluruhan dapat didekati dengan kurva regresi semiparametrik campuran *Spline truncated* linear dan kernel multivariabel. Model regresinya dapat dapat dituliskan seperti pada persamaan (4.11).

$$\begin{aligned} y_i = & f(x_i) + g_1(t_{1i}) + g_2(t_{2i}) + \dots + g_p(t_{pi}) + h_1(z_{1i}) + h_2(z_{2i}) + \dots + h_q(z_{qi}) + \varepsilon_i \\ = & f(x_i) + \sum_{r=1}^p g_r(t_{ri}) + \sum_{s=1}^q h_s(z_{si}) \end{aligned} \quad (4.11)$$

Kurva regresi  $g_r(t_{ri})$  dapat ditulis menjadi :

$$g_r(t_{ri}) = \theta_{r1}x_{ri} + \lambda_{r1}(x_{ri} - K_{r1})_+ + \dots + \lambda_{rm}(x_{ri} - K_{rm})_+ \quad (4.12)$$

dengan  $\theta_{r1}, \lambda_{r1}, \dots, \lambda_{rm}$  merupakan parameter-parameter yang tidak diketahui. untuk  $r = 1$ , diperoleh :

$$g_1(t_{1i}) = \theta_{11}t_{1i} + \lambda_{11}(t_{1i} - K_{11})_+ + \dots + \lambda_{1m}(t_{1i} - K_{1m})_+ \quad (4.13)$$

dengan  $i=1, 2, \dots, n$ , maka menjadi :

$$\begin{pmatrix} g_1(t_{12}) \\ g_1(t_{12}) \\ \vdots \\ g_1(t_{1n}) \end{pmatrix} = \begin{pmatrix} \theta_{11}t_{12} + \lambda_{11}(t_{11} - K_{11})_+ + \dots + \lambda_{1m}(t_{11} - K_{1m})_+ \\ \theta_{11}t_{12} + \lambda_{11}(t_{11} - K_{11})_+ + \dots + \lambda_{1m}(t_{11} - K_{1m})_+ \\ \vdots \\ \theta_{11}t_{1n} + \lambda_{11}(t_{1n} - K_{11})_+ + \dots + \lambda_{1m}(t_{1n} - K_{1m})_+ \end{pmatrix}$$

dapat diubah dalam bentuk matiks menjadi :



$$\begin{pmatrix} g_1(t_{12}) \\ g_1(t_{12}) \\ \vdots \\ g_1(t_{1n}) \end{pmatrix} = \begin{pmatrix} t_{12} & (t_{11} - K_{11})_+ & (t_{11} - K_{1m})_+ \\ t_{12} & (t_{11} - K_{11})_+ & (t_{11} - K_{1m})_+ \\ \vdots & \vdots & \vdots \\ t_{1n} & (t_{1n} - K_{11})_+ & (t_{1n} - K_{1m})_+ \end{pmatrix} \begin{pmatrix} \theta_{11} \\ \lambda_{11} \\ \vdots \\ \lambda_{1m} \end{pmatrix}$$

atau dapat ditulis dalam notasi matriks :

$$\tilde{g}_1(t_1) = \mathbf{G}(k_1)\tilde{\theta}_1 \quad (4.14)$$

Untuk  $r = 2$  dilakukan hal yang sama, sehingga diperoleh  $\tilde{g}_2(t_2) = \mathbf{G}(k_2)\theta_2$ . Jika  $r = 3$  diperoleh  $\tilde{g}_3(t_3) = \mathbf{G}(k_3)\tilde{\theta}_3$ , dan seterusnya sampai pada jika  $r = p$  diperoleh  $\tilde{g}_p(t_p) = \mathbf{G}(k_p)\tilde{\theta}_p$ . Dengan demikian untuk kurva *Spline truncated* linear multivariabel diperoleh :

$$\begin{aligned} \sum_{r=1}^p g_r(t_{ri}) &= g_1(t_{1i}) + g_2(t_{2i}) + \dots + g_p(t_{pi}) \\ &= \mathbf{G}(k_1)\tilde{\theta}_1 + \mathbf{G}(k_2)\tilde{\theta}_2 + \dots + \mathbf{G}(k_p)\tilde{\theta}_p \\ &= \begin{pmatrix} \mathbf{G}(k_1) & \mathbf{G}(k_2) & \dots & \mathbf{G}(k_p) \end{pmatrix} \begin{pmatrix} \tilde{\theta}_1 \\ \tilde{\theta}_2 \\ \vdots \\ \tilde{\theta}_p \end{pmatrix} \\ &= \mathbf{G}(\tilde{k})\tilde{\theta} \end{aligned} \quad (4.15)$$

Sementara itu, kurva regresi  $h_s(z_{si})$  dapat ditulis menjadi :

$$\hat{h}_{\alpha_s}(z_i) = n^{-1} \sum_{j=1}^n W_{\alpha_s}(z_i) y_j, \quad (4.16)$$

untuk  $s=I$ , diperoleh :

$$\hat{h}_{\alpha_1}(z_i) = n^{-1} \sum_{j=1}^n W_{\alpha_1}(z_i) y_j. \quad (4.17)$$

Karena berlaku untuk  $i = 1, 2, \dots, n$ , maka :

$$\begin{pmatrix} \hat{h}_{\alpha_1}(z_1) \\ \hat{h}_{\alpha_1}(z_2) \\ \vdots \\ \hat{h}_{\alpha_1}(z_n) \end{pmatrix} = \begin{pmatrix} n^{-1}W_{\alpha_1}(z_1)y_1 + n^{-1}W_{\alpha_2}(z_1)y_2 + \dots + n^{-1}W_{\alpha_n}(z_1)y_n \\ n^{-1}W_{\alpha_1}(z_2)y_1 + n^{-1}W_{\alpha_2}(z_2)y_2 + \dots + n^{-1}W_{\alpha_n}(z_2)y_n \\ \vdots \\ n^{-1}W_{\alpha_1}(z_n)y_1 + n^{-1}W_{\alpha_2}(z_n)y_2 + \dots + n^{-1}W_{\alpha_n}(z_n)y_n \end{pmatrix}$$

dapat diubah dalam bentuk matiks menjadi :

$$\begin{pmatrix} \hat{h}_{\alpha_1}(z_1) \\ \hat{h}_{\alpha_1}(z_2) \\ \vdots \\ \hat{h}_{\alpha_1}(z_n) \end{pmatrix} = \begin{pmatrix} n^{-1}W_{\alpha_1 1}(z_1) & n^{-1}W_{\alpha_1 2}(z_1) & \cdots & n^{-1}W_{\alpha_1 n}(z_1) \\ n^{-1}W_{\alpha_1 1}(z_2) & n^{-1}W_{\alpha_1 2}(z_2) & \cdots & n^{-1}W_{\alpha_1 n}(z_2) \\ \vdots & \vdots & \vdots & \vdots \\ n^{-1}W_{\alpha_1 1}(z_n) & n^{-1}W_{\alpha_1 2}(z_n) & \cdots & n^{-1}W_{\alpha_1 n}(z_n) \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

atau dapat ditulis dalam notasi matriks :

$$\hat{h}_{\alpha_1}(z) = \mathbf{D}(\alpha_1)\tilde{y} \quad (4.18)$$

Sehingga untuk  $s = 2$  diperoleh  $\hat{h}_{\alpha_2}(z) = \mathbf{D}(\alpha_2)\tilde{y}$ , untuk  $s = 3$  diperoleh

$\hat{h}_{\alpha_3}(z) = \mathbf{D}(\alpha_3)\tilde{y}$ , dan untuk  $s = q$  diperoleh  $\hat{h}_{\alpha_q}(z) = \mathbf{D}(\alpha_q)\tilde{y}$ . Sehingga :

$$\begin{aligned} \sum_{s=1}^q \hat{h}_s(z_s) &= \hat{h}_1(z_1) + \hat{h}_2(z_2) + \dots + \hat{h}_q(z_q) \\ &= \mathbf{D}(\alpha_1)\tilde{y} + \mathbf{D}(\alpha_2)\tilde{y} + \dots + \mathbf{D}(\alpha_q)\tilde{y} \\ &= \mathbf{D}(\tilde{\alpha})\tilde{y}, \end{aligned} \quad (4.19)$$

dengan  $\mathbf{D}(\tilde{\alpha})$  dapat ditulis menjadi :

$$\begin{aligned} \mathbf{D}(\tilde{\alpha}) &= \mathbf{D}(\alpha_1) + \mathbf{D}(\alpha_2) + \dots + \mathbf{D}(\alpha_q) \\ &= \begin{pmatrix} n^{-1}W_{\alpha_1 1}(z_1) & n^{-1}W_{\alpha_1 2}(z_1) & \cdots & n^{-1}W_{\alpha_1 n}(z_1) \\ n^{-1}W_{\alpha_1 1}(z_2) & n^{-1}W_{\alpha_1 2}(z_2) & \cdots & n^{-1}W_{\alpha_1 n}(z_2) \\ \vdots & \vdots & \vdots & \vdots \\ n^{-1}W_{\alpha_1 1}(z_n) & n^{-1}W_{\alpha_1 2}(z_n) & \cdots & n^{-1}W_{\alpha_1 n}(z_n) \end{pmatrix} + \\ &\quad \begin{pmatrix} n^{-1}W_{\alpha_2 1}(z_1) & n^{-1}W_{\alpha_2 2}(z_1) & \cdots & n^{-1}W_{\alpha_2 n}(z_1) \\ n^{-1}W_{\alpha_2 1}(z_2) & n^{-1}W_{\alpha_2 2}(z_2) & \cdots & n^{-1}W_{\alpha_2 n}(z_2) \\ \vdots & \vdots & \vdots & \vdots \\ n^{-1}W_{\alpha_2 1}(z_n) & n^{-1}W_{\alpha_2 2}(z_n) & \cdots & n^{-1}W_{\alpha_2 n}(z_n) \end{pmatrix} + \dots + \\ &\quad \begin{pmatrix} n^{-1}W_{\alpha_q 1}(z_1) & n^{-1}W_{\alpha_q 2}(z_1) & \cdots & n^{-1}W_{\alpha_q n}(z_1) \\ n^{-1}W_{\alpha_q 1}(z_2) & n^{-1}W_{\alpha_q 2}(z_2) & \cdots & n^{-1}W_{\alpha_q n}(z_2) \\ \vdots & \vdots & \vdots & \vdots \\ n^{-1}W_{\alpha_q 1}(z_n) & n^{-1}W_{\alpha_q 2}(z_n) & \cdots & n^{-1}W_{\alpha_q n}(z_n) \end{pmatrix} \end{aligned}$$

$$\mathbf{D}(\tilde{\alpha}) = \begin{pmatrix} n^{-1} \sum_{k=1}^q W_{\alpha_k 1}(z_1) & n^{-1} \sum_{k=1}^q W_{\alpha_k 2}(z_1) & \cdots & n^{-1} \sum_{k=1}^q W_{\alpha_k n}(z_1) \\ n^{-1} \sum_{k=1}^q W_{\alpha_k 1}(z_2) & n^{-1} \sum_{k=1}^q W_{\alpha_k 2}(z_2) & \cdots & n^{-1} \sum_{k=1}^q W_{\alpha_k n}(z_2) \\ \vdots & \vdots & \vdots & \vdots \\ n^{-1} \sum_{k=1}^q W_{\alpha_k 1}(z_n) & n^{-1} \sum_{k=1}^q W_{\alpha_k 2}(z_n) & \cdots & n^{-1} \sum_{k=1}^q W_{\alpha_k n}(z_n) \end{pmatrix}$$

Sehingga model regresi semiparametrik multivariabel campuran *Spline truncated* dan kernel :

$$y_i = f(x_i) + \sum_{r=1}^p g_r(t_{ri}) + \sum_{s=1}^q h_s(z_{si}) + \tilde{\varepsilon} \quad (4.20)$$

dapat ditulis seperti pada persamaan (4.21):

$$\tilde{y} = \mathbf{X}\tilde{\beta} + \mathbf{G}(\tilde{k})\tilde{\theta} + \mathbf{D}(\tilde{\alpha})\tilde{y} + \tilde{\varepsilon} \quad (4.21)$$

## 4.2 Estimasi Kurva Regresi Campuran Semiparametrik *Spline Truncated* dan Kernel

Estimator semiparametrik campuran *Spline truncated* dan kernel dapat diperoleh dengan menggunakan metode *ordinary least square* (OLS). Dari model pada persamaan (4.21) diperoleh :

$$\begin{aligned} \tilde{\varepsilon} &= \tilde{y} - \mathbf{X}\tilde{\beta} - \mathbf{G}(\tilde{k})\tilde{\theta} - \mathbf{D}(\tilde{\alpha})\tilde{y} \\ &= \tilde{y} - \mathbf{D}(\tilde{\alpha})\tilde{y} - \mathbf{X}\tilde{\beta} - \mathbf{G}(\tilde{k})\tilde{\theta} \\ &= (\mathbf{I} - \mathbf{D}(\tilde{\alpha}))\tilde{y} - \mathbf{X}\tilde{\beta} - \mathbf{G}(\tilde{k})\tilde{\theta} \end{aligned}$$

Selanjutnya didapat :

$$\begin{aligned} \sum_{i=1}^n \varepsilon_i^2 &= \tilde{\varepsilon}^T \tilde{\varepsilon} \\ &= \left[ (\mathbf{I} - \mathbf{D}(\tilde{\alpha}))\tilde{y} - \mathbf{X}\tilde{\beta} - \mathbf{G}(\tilde{k})\tilde{\theta} \right]^T \left[ (\mathbf{I} - \mathbf{D}(\tilde{\alpha}))\tilde{y} - \mathbf{X}\tilde{\beta} - \mathbf{G}(\tilde{k})\tilde{\theta} \right] \end{aligned}$$

Kemudian diperoleh :

$$\sum_{i=1}^n \varepsilon_i^2 = \left[ \tilde{y}^T (\mathbf{I} - \mathbf{D}(\tilde{\alpha}))^T - \tilde{\beta}^T \mathbf{X}^T - \tilde{\theta}^T \mathbf{G}(\tilde{k})^T \right] \left[ (\mathbf{I} - \mathbf{D}(\tilde{\alpha}))\tilde{y} - \mathbf{X}\tilde{\beta} - \mathbf{G}(\tilde{k})\tilde{\theta} \right]$$

Jika disederhanakan menjadi :

$$\begin{aligned}\tilde{\varepsilon}^T \tilde{\varepsilon} = & \tilde{y}^T (\mathbf{I} - \mathbf{D}(\tilde{\alpha}))^T (\mathbf{I} - \mathbf{D}(\tilde{\alpha})) \tilde{y} - \tilde{y}^T (\mathbf{I} - \mathbf{D}(\tilde{\alpha}))^T \mathbf{X} \tilde{\beta} - \tilde{y}^T (\mathbf{I} - \mathbf{D}(\tilde{\alpha}))^T \mathbf{G}(\tilde{k}) \tilde{\theta} + \\ & - \tilde{\beta}^T \mathbf{X}^T (\mathbf{I} - \mathbf{D}(\tilde{\alpha})) \tilde{y} + \tilde{\beta}^T \mathbf{X}^T \mathbf{X} \tilde{\beta} + \tilde{\beta}^T \mathbf{X}^T \mathbf{G}(\tilde{k}) \tilde{\theta} - \tilde{\theta}^T \mathbf{G}(\tilde{k})^T (\mathbf{I} - \mathbf{D}(\tilde{\alpha})) \tilde{y} + \\ & \tilde{\theta}^T \mathbf{G}(\tilde{k})^T \mathbf{X} \tilde{\beta} + \tilde{\theta}^T \mathbf{G}(\tilde{k})^T \mathbf{G}(\tilde{k}) \tilde{\theta}\end{aligned}$$

Selanjutnya diperoleh :

$$\begin{aligned}\tilde{\varepsilon}^T \tilde{\varepsilon} = & \tilde{y}^T (\mathbf{I} - \mathbf{D}(\tilde{\alpha}))^T (\mathbf{I} - \mathbf{D}(\tilde{\alpha})) \tilde{y} - \tilde{\beta}^T \mathbf{X}^T (\mathbf{I} - \mathbf{D}(\tilde{\alpha})) \tilde{y} - \tilde{\theta}^T \mathbf{G}(\tilde{k})^T (\mathbf{I} - \mathbf{D}(\tilde{\alpha})) \tilde{y} + \\ & - \tilde{\beta}^T \mathbf{X}^T (\mathbf{I} - \mathbf{D}(\tilde{\alpha})) \tilde{y} + \tilde{\beta}^T \mathbf{X}^T \mathbf{X} \tilde{\beta} + \tilde{\beta}^T \mathbf{X}^T \mathbf{G}(\tilde{k}) \tilde{\theta} - \tilde{\theta}^T \mathbf{G}(\tilde{k})^T (\mathbf{I} - \mathbf{D}(\tilde{\alpha})) \tilde{y} + \\ & \tilde{\beta}^T \mathbf{X}^T \mathbf{G}(\tilde{k}) \tilde{\theta} + \tilde{\theta}^T \mathbf{G}(\tilde{k})^T \mathbf{G}(\tilde{k}) \tilde{\theta}\end{aligned}$$

dengan sedikit penjabaran diperoleh :

$$\begin{aligned}\tilde{\varepsilon}^T \tilde{\varepsilon} = & \tilde{y}^T (\mathbf{I} - \mathbf{D}(\tilde{\alpha}))^T (\mathbf{I} - \mathbf{D}(\tilde{\alpha})) \tilde{y} - 2\tilde{\beta}^T \mathbf{X}^T (\mathbf{I} - \mathbf{D}(\tilde{\alpha})) \tilde{y} + \\ & - 2\tilde{\theta}^T \mathbf{G}(\tilde{k})^T (\mathbf{I} - \mathbf{D}(\tilde{\alpha})) \tilde{y} + \tilde{\beta}^T \mathbf{X}^T \mathbf{X} \tilde{\beta} + 2\tilde{\beta}^T \mathbf{X}^T \mathbf{G}(\tilde{k}) \tilde{\theta} + \\ & \tilde{\theta}^T \mathbf{G}(\tilde{k})^T \mathbf{G}(\tilde{k}) \tilde{\theta} \\ = & Q(\tilde{\beta}, \tilde{\theta})\end{aligned}$$

Selanjutnya Derivatif parsial terhadap  $\tilde{\beta}$  didapat :

$$\frac{\partial Q(\tilde{\beta}, \tilde{\theta})}{\partial \tilde{\beta}} = 0 - 2\mathbf{X}^T (\mathbf{I} - \mathbf{D}(\tilde{\alpha})) \tilde{y} + 2\mathbf{X}^T \mathbf{X} \hat{\beta} + 2\tilde{\beta}^T \mathbf{X}^T \mathbf{G}(\tilde{k}) \tilde{\theta}$$

Sementara itu Derivatif parsial terhadap  $\tilde{\theta}$  didapat :

$$\frac{\partial Q(\tilde{\beta}, \tilde{\theta})}{\partial \tilde{\theta}} = 0 - 2\mathbf{G}(\tilde{k})^T (\mathbf{I} - \mathbf{D}(\tilde{\alpha})) \tilde{y} + 2\mathbf{G}(\tilde{k})^T \mathbf{X} \tilde{\beta} + 2\mathbf{G}(\tilde{k})^T \mathbf{G}(\tilde{k}) \tilde{\theta}$$

Selanjutnya derivatif-derivatif tersebut disamakan dengan nol, yaitu :

$$\frac{\partial Q(\tilde{\beta}, \tilde{\theta})}{\partial \tilde{\beta}} = 0,$$

didapat persamaan :

$$-2\mathbf{X}^T (\mathbf{I} - \mathbf{D}(\tilde{\alpha})) \tilde{y} + 2\mathbf{X}^T \mathbf{X} \hat{\beta} + 2\mathbf{X}^T \mathbf{G}(\tilde{k}) \tilde{\theta} = 0$$

Selanjutnya dapat ditulis :

$$\mathbf{X}^T \mathbf{X} \hat{\beta} = \mathbf{X}^T (\mathbf{I} - \mathbf{D}(\tilde{\alpha})) \tilde{y} - \mathbf{X}^T \mathbf{G}(\tilde{k}) \tilde{\theta}$$

Sehingga diperoleh :

$$\begin{aligned}\hat{\beta} &= (\mathbf{X}^T \mathbf{X})^{-1} (\mathbf{X}^T (\mathbf{I} - \mathbf{D}(\tilde{\alpha})) \tilde{\mathbf{y}} - \mathbf{X}^T \mathbf{G}(\tilde{k}) \tilde{\theta}) \\ &= (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T (\mathbf{I} - \mathbf{D}(\tilde{\alpha})) \tilde{\mathbf{y}} - (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{G}(\tilde{k}) \tilde{\theta}\end{aligned}\quad (4.22)$$

Selanjutnya untuk :

$$\frac{\partial Q(\tilde{\beta}, \tilde{\theta})}{\partial \tilde{\theta}} = 0$$

didapat persamaan :

$$-2\mathbf{G}(\tilde{k})^T (\mathbf{I} - \mathbf{D}(\tilde{\alpha})) \tilde{\mathbf{y}} + 2\mathbf{G}(\tilde{k})^T \mathbf{X} \tilde{\beta} + 2\mathbf{G}(\tilde{k})^T \mathbf{G}(\tilde{k}) \hat{\tilde{\theta}} = 0$$

Selanjutnya dapat ditulis :

$$\mathbf{G}(\tilde{k})^T \mathbf{G}(\tilde{k}) \hat{\tilde{\theta}} = \mathbf{G}(\tilde{k})^T (\mathbf{I} - \mathbf{D}(\tilde{\alpha})) \tilde{\mathbf{y}} - \mathbf{G}(\tilde{k})^T \mathbf{X} \tilde{\beta}$$

Sehingga diperoleh :

$$\begin{aligned}\hat{\tilde{\theta}} &= (\mathbf{G}(\tilde{k})^T \mathbf{G}(\tilde{k}))^{-1} (\mathbf{G}(\tilde{k})^T (\mathbf{I} - \mathbf{D}(\tilde{\alpha})) \tilde{\mathbf{y}} - \mathbf{G}(\tilde{k})^T \mathbf{X} \tilde{\beta}) \\ &= (\mathbf{G}(\tilde{k})^T \mathbf{G}(\tilde{k}))^{-1} \mathbf{G}(\tilde{k})^T (\mathbf{I} - \mathbf{D}(\tilde{\alpha})) \tilde{\mathbf{y}} - (\mathbf{G}(\tilde{k})^T \mathbf{G}(\tilde{k}))^{-1} \mathbf{G}(\tilde{k})^T \mathbf{X} \tilde{\beta}\end{aligned}\quad (4.23)$$

Kemudian mensubstitusikan  $\hat{\tilde{\theta}}$  ke dalam persamaan (4.22), didapat :

$$\begin{aligned}\hat{\beta} &= \left[ (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T (\mathbf{I} - \mathbf{D}(\tilde{\alpha})) \tilde{\mathbf{y}} - (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{G}(\tilde{k}) \right] \left[ (\mathbf{G}(\tilde{k})^T \mathbf{G}(\tilde{k}))^{-1} \mathbf{G}(\tilde{k})^T (\mathbf{I} - \mathbf{D}(\tilde{\alpha})) \tilde{\mathbf{y}} + \right. \\ &\quad \left. - (\mathbf{G}(\tilde{k})^T \mathbf{G}(\tilde{k}))^{-1} \mathbf{G}(\tilde{k})^T \mathbf{X} \tilde{\beta} \right] \\ &= (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T (\mathbf{I} - \mathbf{D}(\tilde{\alpha})) \tilde{\mathbf{y}} - (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{G}(\tilde{k}) (\mathbf{G}(\tilde{k})^T \mathbf{G}(\tilde{k}))^{-1} \mathbf{G}(\tilde{k})^T (\mathbf{I} - \mathbf{D}(\tilde{\alpha})) \tilde{\mathbf{y}} + \\ &\quad (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{G}(\tilde{k}) (\mathbf{G}(\tilde{k})^T \mathbf{G}(\tilde{k}))^{-1} \mathbf{G}(\tilde{k})^T \mathbf{X} \tilde{\beta}\end{aligned}$$

dengan menyederhanakan didapat :

$$\begin{aligned}\hat{\beta} &- (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{G}(\tilde{k}) (\mathbf{G}(\tilde{k})^T \mathbf{G}(\tilde{k}))^{-1} \mathbf{G}(\tilde{k})^T \mathbf{X} \tilde{\beta} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T (\mathbf{I} - \mathbf{D}(\tilde{\alpha})) \tilde{\mathbf{y}} + \\ &- (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{G}(\tilde{k}) (\mathbf{G}(\tilde{k})^T \mathbf{G}(\tilde{k}))^{-1} \mathbf{G}(\tilde{k})^T (\mathbf{I} - \mathbf{D}(\tilde{\alpha})) \tilde{\mathbf{y}}\end{aligned}$$

Selanjutnya dapat ditulis menjadi :

$$\begin{aligned}\left[ \mathbf{I} - (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{G}(\tilde{k}) (\mathbf{G}(\tilde{k})^T \mathbf{G}(\tilde{k}))^{-1} \mathbf{G}(\tilde{k})^T \mathbf{X} \right] \hat{\beta} &= \left[ (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T (\mathbf{I} - \mathbf{D}(\tilde{\alpha})) + \right. \\ &- \left. (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{G}(\tilde{k}) (\mathbf{G}(\tilde{k})^T \mathbf{G}(\tilde{k}))^{-1} \mathbf{G}(\tilde{k})^T (\mathbf{I} - \mathbf{D}(\tilde{\alpha})) \right] \tilde{\mathbf{y}}\end{aligned}$$

Kemudian diperoleh :

$$\hat{\beta} = \left[ \mathbf{I} - (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{G}(\tilde{k}) (\mathbf{G}(\tilde{k})^T \mathbf{G}(\tilde{k}))^{-1} \mathbf{G}(\tilde{k})^T \mathbf{X} \right]^{-1} \left[ (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T (\mathbf{I} - \mathbf{D}(\tilde{\alpha})) + \right. \\ \left. - (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{G}(\tilde{k}) (\mathbf{G}(\tilde{k})^T \mathbf{G}(\tilde{k}))^{-1} \mathbf{G}(\tilde{k})^T (\mathbf{I} - \mathbf{D}(\tilde{\alpha})) \right] \tilde{y}$$

Estimator  $\hat{\beta}$  dapat ditulis menjadi :

$$\hat{\beta} = \mathbf{A}(\tilde{k}, \tilde{\alpha}) \tilde{y}, \quad (4.24)$$

dimana :

$$\mathbf{A}(\tilde{k}, \tilde{\alpha}) = \left[ \mathbf{I} - (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{G}(\tilde{k}) (\mathbf{G}(\tilde{k})^T \mathbf{G}(\tilde{k}))^{-1} \mathbf{G}(\tilde{k})^T \mathbf{X} \right]^{-1} \left[ (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T (\mathbf{I} - \mathbf{D}(\tilde{\alpha})) + \right. \\ \left. - (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{G}(\tilde{k}) (\mathbf{G}(\tilde{k})^T \mathbf{G}(\tilde{k}))^{-1} \mathbf{G}(\tilde{k})^T (\mathbf{I} - \mathbf{D}(\tilde{\alpha})) \right]$$

Selanjutnya, substitusikan  $\tilde{\beta}$  ke dalam persamaan (4.23) didapat :

$$\hat{\theta} = (\mathbf{G}(\tilde{k})^T \mathbf{G}(\tilde{k}))^{-1} \mathbf{G}(\tilde{k})^T (\mathbf{I} - \mathbf{D}(\tilde{\alpha})) \tilde{y} - (\mathbf{G}(\tilde{k})^T \mathbf{G}(\tilde{k}))^{-1} \mathbf{G}(\tilde{k})^T \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T (\mathbf{I} - \mathbf{D}(\tilde{\alpha})) \tilde{y} + \\ - (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{G}(\tilde{k}) \tilde{\theta}$$

Kemudian diperoleh :

$$\hat{\theta} = (\mathbf{G}(\tilde{k})^T \mathbf{G}(\tilde{k}))^{-1} \mathbf{G}(\tilde{k})^T (\mathbf{I} - \mathbf{D}(\tilde{\alpha})) \tilde{y} - (\mathbf{G}(\tilde{k})^T \mathbf{G}(\tilde{k}))^{-1} \mathbf{G}(\tilde{k})^T \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T (\mathbf{I} - \mathbf{D}(\tilde{\alpha})) \tilde{y} + \\ (\mathbf{G}(\tilde{k})^T \mathbf{G}(\tilde{k}))^{-1} \mathbf{G}(\tilde{k})^T \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{G}(\tilde{k}) \tilde{\theta}$$

dengan sedikit penjabaran didapat :

$$\hat{\theta} - (\mathbf{G}(\tilde{k})^T \mathbf{G}(\tilde{k}))^{-1} \mathbf{G}(\tilde{k})^T \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{G}(\tilde{k}) \tilde{\theta} = (\mathbf{G}(\tilde{k})^T \mathbf{G}(\tilde{k}))^{-1} \mathbf{G}(\tilde{k})^T (\mathbf{I} - \mathbf{D}(\tilde{\alpha})) \tilde{y} + \\ - (\mathbf{G}(\tilde{k})^T \mathbf{G}(\tilde{k}))^{-1} \mathbf{G}(\tilde{k})^T \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T (\mathbf{I} - \mathbf{D}(\tilde{\alpha})) \tilde{y}$$

Persamaan diatas dapat diuraikan menjadi :

$$\left( \mathbf{I} - (\mathbf{G}(\tilde{k})^T \mathbf{G}(\tilde{k}))^{-1} \mathbf{G}(\tilde{k})^T \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{G}(\tilde{k}) \right) \hat{\theta} = (\mathbf{G}(\tilde{k})^T \mathbf{G}(\tilde{k}))^{-1} \mathbf{G}(\tilde{k})^T (\mathbf{I} - \mathbf{D}(\tilde{\alpha})) \tilde{y} + \\ - (\mathbf{G}(\tilde{k})^T \mathbf{G}(\tilde{k}))^{-1} \mathbf{G}(\tilde{k})^T \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T (\mathbf{I} - \mathbf{D}(\tilde{\alpha})) \tilde{y}$$

Selanjutnya diperoleh :

$$\hat{\theta} = \left[ \mathbf{I} - (\mathbf{G}(\tilde{k})^T \mathbf{G}(\tilde{k}))^{-1} \mathbf{G}(\tilde{k})^T \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{G}(\tilde{k}) \right]^{-1} (\mathbf{G}(\tilde{k})^T \mathbf{G}(\tilde{k}))^{-1} \mathbf{G}(\tilde{k})^T (\mathbf{I} - \mathbf{D}(\tilde{\alpha})) \tilde{y} + \\ - (\mathbf{G}(\tilde{k})^T \mathbf{G}(\tilde{k}))^{-1} \mathbf{G}(\tilde{k})^T \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T (\mathbf{I} - \mathbf{D}(\tilde{\alpha})) \tilde{y}$$

Estimator  $\hat{\theta}$  dapat ditulis menjadi :

$$\hat{\theta} = \mathbf{B}(\tilde{k}, \tilde{\alpha}) \tilde{y}, \quad (4.25)$$

dimana :

$$\mathbf{B}(\tilde{k}, \tilde{\alpha}) = \left[ \mathbf{I} - \left( \mathbf{G}(\tilde{k})^T \mathbf{G}(\tilde{k}) \right)^{-1} \mathbf{G}(\tilde{k})^T \mathbf{X} \left( \mathbf{X}^T \mathbf{X} \right)^{-1} \mathbf{X}^T \mathbf{G}(\tilde{k}) \right]^{-1} \left( \mathbf{G}(\tilde{k})^T \mathbf{G}(\tilde{k}) \right)^{-1} \mathbf{G}(\tilde{k})^T (\mathbf{I} - \mathbf{D}(\tilde{\alpha})) + \left( \mathbf{G}(\tilde{k})^T \mathbf{G}(\tilde{k}) \right)^{-1} \mathbf{G}(\tilde{k})^T \mathbf{X} \left( \mathbf{X}^T \mathbf{X} \right)^{-1} \mathbf{X}^T (\mathbf{I} - \mathbf{D}(\tilde{\alpha})) \right]$$

Berdasarkan  $\hat{\beta}$  didapat estimator kurva parametrik linear :

$$\begin{aligned} \hat{f}(x) &= \mathbf{X} \hat{\beta} \\ &= \mathbf{X} \left[ \mathbf{I} - \left( \mathbf{X}^T \mathbf{X} \right)^{-1} \mathbf{X}^T \mathbf{G}(\tilde{k}) \left( \mathbf{G}(\tilde{k})^T \mathbf{G}(\tilde{k}) \right)^{-1} \mathbf{G}(\tilde{k})^T \mathbf{X} \right]^{-1} \left[ \left( \mathbf{X}^T \mathbf{X} \right)^{-1} \mathbf{X}^T (\mathbf{I} - \mathbf{D}(\tilde{\alpha})) + \right. \\ &\quad \left. - \left( \mathbf{X}^T \mathbf{X} \right)^{-1} \mathbf{X}^T \mathbf{G}(\tilde{k}) \left( \mathbf{G}(\tilde{k})^T \mathbf{G}(\tilde{k}) \right)^{-1} \mathbf{G}(\tilde{k})^T (\mathbf{I} - \mathbf{D}(\tilde{\alpha})) \right] \tilde{y} \\ &= \mathbf{C}(\tilde{k}, \tilde{\alpha}) \tilde{y} \end{aligned} \quad (4.26)$$

dimana :

$$\mathbf{C}(\tilde{k}, \tilde{\alpha}) = \mathbf{X} \left[ \mathbf{I} - \left( \mathbf{X}^T \mathbf{X} \right)^{-1} \mathbf{X}^T \mathbf{G}(\tilde{k}) \left( \mathbf{G}(\tilde{k})^T \mathbf{G}(\tilde{k}) \right)^{-1} \mathbf{G}(\tilde{k})^T \mathbf{X} \right]^{-1} \left[ \left( \mathbf{X}^T \mathbf{X} \right)^{-1} \mathbf{X}^T (\mathbf{I} - \mathbf{D}(\tilde{\alpha})) + \right. \\ \left. - \left( \mathbf{X}^T \mathbf{X} \right)^{-1} \mathbf{X}^T \mathbf{G}(\tilde{k}) \left( \mathbf{G}(\tilde{k})^T \mathbf{G}(\tilde{k}) \right)^{-1} \mathbf{G}(\tilde{k})^T (\mathbf{I} - \mathbf{D}(\tilde{\alpha})) \right]$$

Berdasarkan  $\hat{\theta}$  didapat estimator kurva regresi *Spline truncated* linear :

$$\begin{aligned} \tilde{g}(t) &= \mathbf{G}(\tilde{k}) \hat{\theta} \\ &= \mathbf{G}(\tilde{k}) \left[ \mathbf{I} - \left( \mathbf{G}(\tilde{k})^T \mathbf{G}(\tilde{k}) \right)^{-1} \mathbf{G}(\tilde{k})^T \mathbf{X} \left( \mathbf{X}^T \mathbf{X} \right)^{-1} \mathbf{X}^T \mathbf{G}(\tilde{k}) \right]^{-1} \left( \mathbf{G}(\tilde{k})^T \mathbf{G}(\tilde{k}) \right)^{-1} \mathbf{G}(\tilde{k})^T (\mathbf{I} - \mathbf{D}(\tilde{\alpha})) + \right. \\ &\quad \left. - \left( \mathbf{X}^T \mathbf{X} \right)^{-1} \mathbf{X}^T \mathbf{G}(\tilde{k}) \left( \mathbf{G}(\tilde{k})^T \mathbf{G}(\tilde{k}) \right)^{-1} \mathbf{G}(\tilde{k})^T (\mathbf{I} - \mathbf{D}(\tilde{\alpha})) \right] \tilde{y} \\ &= \mathbf{K}(\tilde{k}, \tilde{\alpha}) \tilde{y} \end{aligned} \quad (4.27)$$

dimana :

$$\mathbf{K}(\tilde{k}, \tilde{\alpha}) = \mathbf{G}(\tilde{k}) \left[ \mathbf{I} - \left( \mathbf{G}(\tilde{k})^T \mathbf{G}(\tilde{k}) \right)^{-1} \mathbf{G}(\tilde{k})^T \mathbf{X} \left( \mathbf{X}^T \mathbf{X} \right)^{-1} \mathbf{X}^T \mathbf{G}(\tilde{k}) \right]^{-1} \left( \mathbf{G}(\tilde{k})^T \mathbf{G}(\tilde{k}) \right)^{-1} \mathbf{G}(\tilde{k})^T (\mathbf{I} - \mathbf{D}(\tilde{\alpha})) + \right. \\ \left. - \left( \mathbf{X}^T \mathbf{X} \right)^{-1} \mathbf{X}^T \mathbf{G}(\tilde{k}) \left( \mathbf{G}(\tilde{k})^T \mathbf{G}(\tilde{k}) \right)^{-1} \mathbf{G}(\tilde{k})^T (\mathbf{I} - \mathbf{D}(\tilde{\alpha})) \right].$$

Estimator campuran *Spline truncated* linear dan kernel dalam regresi semiparametrik diperoleh :

$$\begin{aligned}
\hat{\mu}(x, t, z) &= \hat{f}(x) + \hat{g}(t) + \hat{h}(z) \\
&= \mathbf{C}(\tilde{k}, \tilde{\alpha})\tilde{y} + \mathbf{K}(\tilde{k}, \tilde{\alpha})\tilde{y} + \mathbf{D}(\tilde{\alpha})\tilde{y} \\
&= (\mathbf{C}(\tilde{k}, \tilde{\alpha}) + \mathbf{K}(\tilde{k}, \tilde{\alpha}) + \mathbf{D}(\tilde{\alpha}))\tilde{y} \\
&= \mathbf{M}(\tilde{k}, \tilde{\alpha})\tilde{y},
\end{aligned} \tag{4.28}$$

dimana :

$$\mathbf{M}(\tilde{k}, \tilde{\alpha}) = \mathbf{C}(\tilde{k}, \tilde{\alpha}) + \mathbf{K}(\tilde{k}, \tilde{\alpha}) + \mathbf{D}(\tilde{\alpha}) \tag{4.29}$$

Selanjutnya, pemilihan titik knok  $k$  optimal dan *bandwidth*  $\tilde{\alpha}$  optimal menggunakan metode GCV. Sehingga nilai GCV dari model campuran *Spline truncated* dan Kernel diperoleh dengan cara :

$$GCV(k, \alpha) = \frac{MSE(k, \alpha)}{(n^{-1}tr(\mathbf{I} - \mathbf{M}(k, \alpha)))^2} \tag{4.30}$$

dimana :

$$MSE(k, \alpha) = n^{-1} \sum_{i=1}^n (y_i - \hat{y}_i)^2, \text{ dan matriks } \mathbf{M}(k, \alpha) \text{ diperoleh dari persamaan (4.29).}$$

### 4.3 Aplikasi pada Data Produksi Padi di Jawa Tengah

Berikut ini akan dibahas penerapan dari model regresi semiparametrik campuran *Spline truncated* linear dan kernel pada data produksi padi di Jawa Tengah tahun 2008.

#### 4.3.1 Analisis Deskriptif Data

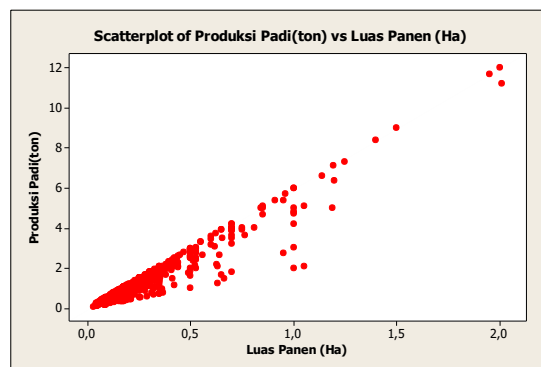
Data yang digunakan merupakan data produksi padi, dengan variabel responnya adalah produksi padi, sementara itu variabel prediktornya adalah luas panen ( $x_1$ ), penggunaan pupuk ( $x_2$ ), penggunaan pestisida ( $x_3$ ), penggunaan benih ( $x_4$ ), dan penggunaan tenaga kerja ( $x_5$ ). Adapun unit observasinya adalah sejumlah 691 rumah tangga tani padi di Jawa Tengah. Berikut deskriptif data yang digunakan :



Tabel 4.1 Deskriptif Data

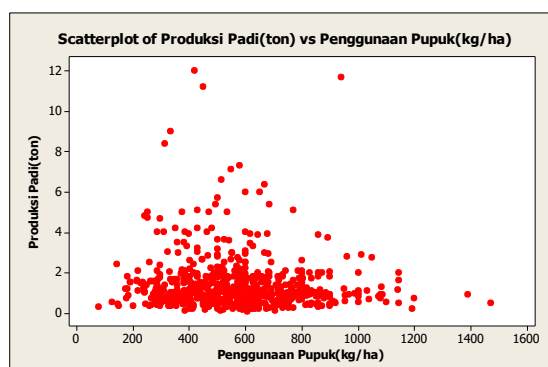
Variabel	Range	Minimum	Maksimum	Jumlah	Rata-Rata	Variansi
Produksi Padi	11,93	0,07	12,00	946,81	1,37	1,72
Luas Panen	1,98	0,03	2,01	192,24	0,28	0,06
Penggunaan Pupuk	1393,67	76,92	1470,59	388499,87	562,23	41758,13
Penggunaan Pestisida	12,87	0,03	12,90	1398,60	2,02	5,36
Penggunaan Benih	65	15,00	80	36183,43	52,36	218,50
Penggunaan Tenaga Kerja	555	10	565	126379	182,8929	9767,983

Selanjutnya untuk melihat bentuk pola hubungan antara variabel respon dengan masing-masing variabel prediktor dapat dilihat dari *scatter plot* berikut :



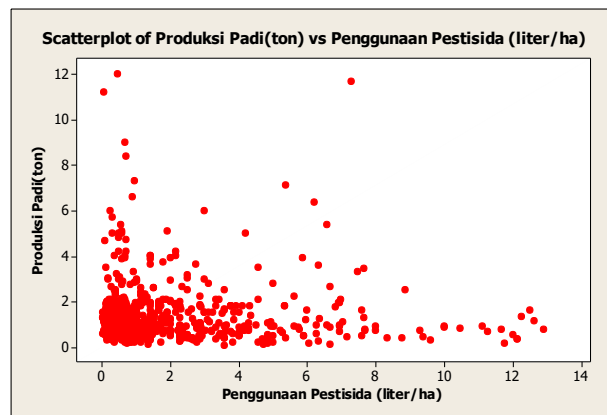
Gambar 4.1 *Scatter plot* antara Produksi Padi dengan Luas Panen

Berdasarkan Gambar 4.1 terlihat bahwa bentuk pola hubungan antara variabel respon produksi padi ( $y$ ) dan variabel prediktor luas panen ( $x_1$ ) cenderung linear, sehingga dicoba dimodelkan menggunakan fungsi parametrik linear.



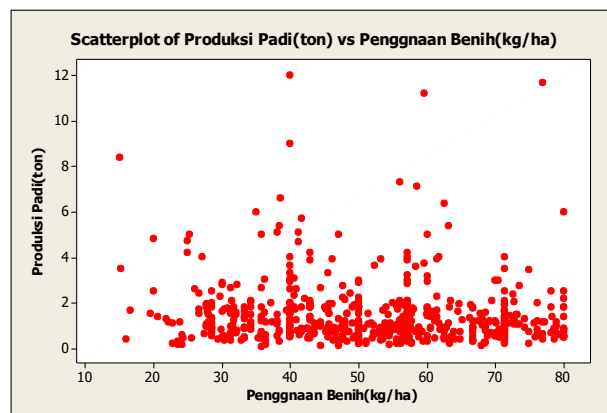
Gambar 4.2 *Scatter plot* antara Produksi Padi dengan Penggunaan Pupuk

Gambar 4.2 menunjukkan bahwa bentuk pola hubungan antara variabel respon produksi padi ( $y$ ) dan variabel prediktor penggunaan pupuk ( $x_2$ ) tidak mempunyai bentuk pola hubungan, sehingga dapat dimodelkan secara nonparametrik. Bentuk pola hubungan antara produksi padi ( $y$ ) dengan penggunaan pupuk ( $x_2$ ) cenderung mengalami perubahan perilaku pada titik sekitar 600. Terlihat bahwa pola data pada interval sebelum 600 cenderung naik sementara pada interval setelah 600 cenderung turun, sehingga dapat dimodelkan dengan fungsi *Spline truncated*.



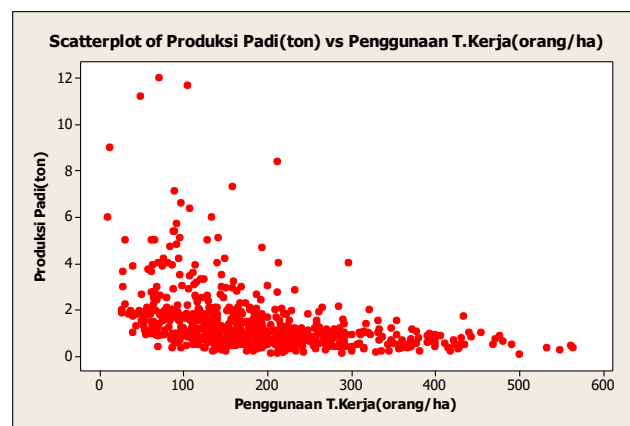
Gambar 4.3 *Scatter plot* antara Produksi Padi dengan Penggunaan Pesticida

Bentuk pola hubungan antara variabel respon produksi padi ( $y$ ) dan variabel prediktor penggunaan pestisida ( $x_3$ ) yang terlihat dari Gambar 4.3 cenderung tidak mengikuti pola tertentu, sehingga dimodelkan secara nonparametrik kernel.



Gambar 4.4 *Scatter plot* antara Produksi Padi dengan Penggunaan Benih

Gambar 4.4 memperlihatkan bahwa bentuk pola hubungan antara variabel respon produksi padi ( $y$ ) dan variabel prediktor penggunaan benih ( $x_4$ ) tidak diketahui bentuk pola hubungannya, sehingga dimodelkan secara nonparametrik. Bentuk pola hubungan antara produksi padi ( $y$ ) dengan penggunaan benih ( $x_4$ ) cenderung mengalami perubahan perilaku pada titik sekitaran 40 dan 60, dimana terlihat perbedaan pola kecendrungan data pada interval sebelum titik 40, interval antara titik 40 sampai titik 60 serta interval setelah titik 60. Sehingga secara teori dapat didekati dengan fungsi *Spline truncated*.



Gambar 4.5 *Scatter plot* antara Produksi Padi dengan Penggunaan Tenaga Kerja

Bentuk pola hubungan antara variabel respon produksi padi ( $y$ ) dan variabel prediktor penggunaan tenaga kerja ( $x_5$ ) tidak diketahui bentuk pola hubungannya, sehingga dimodelkan secara nonparametrik. Bentuk pola hubungan antara produksi padi ( $y$ ) dengan penggunaan tenaga kerja ( $x_5$ ) yang ditunjukkan oleh Gambar 4.5 cenderung mengalami perubahan perilaku pada titik sekitar 100. Terlihat bahwa pola data pada interval sebelum 100 cenderung naik, sementara pada interval setelah 600 cenderung turun. Sehingga secara teori dapat didekati dengan fungsi *Spline truncated*.

Selain melihat bentuk pola hubungan data melalui *scatter plot*, untuk menentukan variabel mana yang didekati dengan *Spline truncated* atau kernel untuk variabel-variabel nonparametrik, juga dilakukan dengan memeriksa nilai GCV dari masing-masing variabel prediktor menggunakan persamaan (4.30). Berdasarkan hasil perhitungan diperoleh nilai GCV untuk semua kemungkinan

model campuran *Spline truncated* dan Kernel seperti terlihat pada Tabel 4.2, Tabel 4.3 dan Tabel 4.4.

Tabel 4.2 GCV dari Model-Model dengan 1 *Spline* dan 3 Kernel

No	Variabel		GCV
	<i>Spline</i>	Kernel	
1	$x_2$	$x_3, x_4, x_5$	0,00031328792
2	$x_3$	$x_2, x_4, x_5$	0,00031327406
3	$x_4$	$x_2, x_3, x_5$	0,00038182276
4	$x_5$	$x_2, x_3, x_4$	0,00025154762*

Berdasarkan Tabel 4.2, dicobakan untuk komponen *Spline* terdiri dari 1 variabel dan komponen kernel terdiri dari 3 variabel. Dari semua kemungkinan diperoleh GCV minimum 0,00025154762 jika komponen *Splinenya* adalah  $x_5$  dan komponen kernelnya adalah  $x_2, x_3, x_4$ .

Tabel 4.3 GCV dari Model-Model dengan 2 *Spline* dan 2 Kernel

No	Variabel		GCV
	<i>Spline</i>	Kernel	
1	$x_2, x_3$	$x_4, x_5$	0,00047521137
2	$x_2, x_4$	$x_3, x_5$	0,00048107085
3	$x_2, x_5$	$x_3, x_4$	0,00022037370
4	$x_3, x_4$	$x_2, x_5$	0,0002203538
5	$x_3, x_5$	$x_2, x_4$	0,00021695737*
6	$x_4, x_5$	$x_2, x_3$	0,00022221830

Berdasarkan Tabel 4.3, dicobakan untuk komponen *Spline* terdiri dari 2 variabel dan komponen kernel terdiri dari 2 variabel. Dari semua kemungkinan diperoleh GCV minimum 0,00021695737 jika komponen *Splinenya* adalah  $x_3, x_5$  dan komponen kernelnya adalah  $x_2, x_4$ .

Tabel 4.4 GCV dari Model-Model dengan 3 *Spline* dan 1 Kernel

No	Variabel		GCV
	<i>Spline</i>	Kernel	
1	$x_3, x_4, x_5$	$x_2$	0,0020932217
2	$x_2, x_4, x_5$	$x_3$	0,00021272675*
3	$x_2, x_3, x_5$	$x_4$	0,00021348236
4	$x_2, x_3, x_4$	$x_5$	0,0019302908

Berdasarkan Tabel 4.4, dicobakan untuk komponen *Spline* terdiri dari 3 variabel dan komponen kernel terdiri dari 1 variabel. Dari semua kemungkinan diperoleh GCV minimum jika komponen *Splinenya* adalah  $x_2, x_4, x_5$  dan komponen kernelnya adalah  $x_3$  dengan nilai GCV sebesar 0,00021695737 yang merupakan nilai GCV terkecil diantara semua kemungkinan yang dicobakan pada Tabel 4.2, 4.3 dan 4.4. Dengan demikian data produksi padi didekati dengan regresi semiparametrik campuran *Spline truncated* dan kernel, dimana terdapat 1 variabel prediktor yang didekati dengan parametrik yaitu variabel luas panen, sementara keempat variabel lainnya didekati dengan nonparametrik. Pada komponen nonparametrik, variabel penggunaan pupuk, penggunaan benih dan penggunaan tenaga kerja didekati dengan fungsi *Spline truncated*, sementara itu, variabel penggunaan pestisida didekati dengan fungsi kernel. Selanjutnya variabel yang didekati dengan parametrik linear disimbolkan dengan  $x$ , variabel yang didekati dengan fungsi *Spline truncated* disimbolkan dengan  $t$  dan variabel yang didekati dengan fungsi kernel disimbolkan dengan  $z$ . Daftar keterangan lengkap dari masing-masing variabel prediktor dapat dilihat pada Tabel 4.5.

Tabel 4.5 Komponen Parametrik dan Nonparametrik

No	Variabel Prediktor	Kurva Regresi	Pendekatan yang digunakan	Simbol Variabel
1	Luas Panen	Parametrik	Parametrik Linear	$x$
2	Penggunaan pupuk	Nonparametrik	Fungsi <i>Spline Truncated</i>	$t_1$
3	Penggunaan benih		Fungsi <i>Spline Truncated</i>	$t_2$
4	Penggunaan Tenaga Kerja		Fungsi <i>Spline Truncated</i>	$t_3$
5	Penggunaan pestisida		Fungsi Kernel	$z$

Berdasarkan Tabel 4.5 terlihat bahwa variabel prediktor luas panen yang semula disimbolkan dengan  $x_1$  menjadi  $x$ . Sementara itu, variabel prediktor penggunaan pupuk yang semula disimbolkan  $x_2$  menjadi  $t_1$ , variabel prediktor penggunaan pestisida yang semula disimbolkan  $x_3$  menjadi  $z$ , variabel prediktor penggunaan benih yang semula disimbolkan  $x_4$  menjadi  $t_2$ , dan variabel prediktor penggunaan tenaga kerja yang semula disimbolkan  $x_5$  menjadi  $t_3$ .

#### 4.3.2 Model Regresi Semiparametrik Campuran

Pola data berdasarkan pada *scatter plot* menunjukkan bahwa terdapat satu variabel yang didekati dengan parametrik linear, tiga variabel yang didekati dengan fungsi *Spline truncated* dan satu variabel yang didekati dengan fungsi kernel, sehingga model regresi semiparametrik campuran *Spline truncated* dan kernel pada data produksi padi dengan  $k$  titik knot dapat ditulis sebagai berikut :

$$\begin{aligned} y_i &= \mu(x_i, t_{1i}, t_{2i}, t_{3i}, z_i) + \varepsilon_i \\ &= f(x_i) + g_1(t_{1i}) + g_2(t_{2i}) + g_3(t_{3i}) + h(z_i) + \varepsilon_i \end{aligned}$$

Kemudian  $f(x_i)$ ,  $g_1(t_{1i})$ ,  $g_2(t_{2i})$ ,  $g_3(t_{3i})$  dan  $h(z_i)$  dihampiri dengan fungsi sebagai berikut :

$$\begin{aligned} y_i &= \beta_0 + \beta_1 x_i + \theta_{11} t_{1i} + \lambda_{11} (t_{1i} - K_{11})_+ + \lambda_{12} (t_{1i} - K_{12})_+ + \dots + \lambda_{1k} (t_{1i} - K_{1k})_+ \\ &\quad + \theta_{21} t_{2i} + \lambda_{21} (t_{2i} - K_{21})_+ + \lambda_{22} (t_{2i} - K_{22})_+ + \dots + \lambda_{2k} (t_{2i} - K_{2k})_+ + \\ &\quad + \theta_{31} t_{3i} + \lambda_{31} (t_{3i} - K_{31})_+ + \lambda_{32} (t_{3i} - K_{32})_+ + \dots + \lambda_{3k} (t_{3i} - K_{3k})_+ + \\ &\quad + \sum_{i=1}^n \frac{\frac{1}{\alpha} K\left(\frac{z - z_i}{\alpha}\right)}{\sum_{i=1}^n \frac{1}{\alpha} K\left(\frac{z - z_i}{\alpha}\right)} Y_i + \varepsilon_i \end{aligned}$$

Pemilihan model regresi semiparametrik campuran *Spline truncated* dan kernel terbaik diperoleh dari penentuan titik-titik knot optimal dan *bandwidth* optimal yang didapatkan dari membandingkan nilai-nilai GCV yang diperoleh. Pada penelitian ini dicobakan untuk 1 titik knot, 2 titik knot dan 3 titik knot, dengan model terbaik yaitu yang mempunyai nilai GCV terkecil.

#### 4.3.2.1 Model dengan Komponen *Spline Truncated* 1 Titik Knot

Berikut ini model regresi semiparametrik campuran *Spline truncated* linear dan kernel dengan masing-masing 1 titik knot pada komponen *Spline truncated*.

$$y_i = \beta_0 + \beta_1 x_i + \theta_{11} t_{1i} + \lambda_{11} (t_{1i} - K_{11})_+ + \theta_{21} t_{2i} + \lambda_{21} (t_{2i} - K_{21})_+ + \theta_{31} t_{3i} + \lambda_{31} (t_{3i} - K_{31})_+ + \sum_{i=1}^n \frac{1}{\alpha} K\left(\frac{z - z_i}{\alpha}\right) Y_i + \varepsilon_i$$

Nilai GCV yang dihasilkan dapat dilihat pada Tabel 4.6.

Tabel 4.6 Perbandingan nilai GCV Model Komponen *Spline* dengan 1 Titik Knot

Parametrik Linear	<i>Spline</i>			Kernel	$GCV = (K_1, K_2, K_3, \alpha)$
Beta 1	Knot (K)			Bandwidth ( $\alpha$ )	
$x$	$t_1 = K_1$	$t_2 = K_1$	$t_3 = K_1$	$\alpha$	
5,236	143,285	18,095	36,429	2,948	0,00021272675*
5,257	209,650	21,190	62,857	2,969	0,00021322790
5,287	276,016	24,286	89,286	2,942	0,00021484087
5,331	674,207	42,857	247,857	3,025	0,00021579064
5,330	607,842	39,762	221,429	3,009	0,00021588836

Berdasarkan Tabel 4.6 didapat GCV terkecil 0,00021272675. Berdasarkan GCV terkecil ini memberikan estimasi parameter-parameter pada komponen *Spline* disajikan pada Tabel 4.7 :

Tabel 4.7 Estimasi Parameter Komponen *Spline Truncated* dengan 1 Titik Knot

Variabel	Parameter	Estimasi
$t_1$	$\theta_{11}$	0,0034
	$\lambda_{11}$	-0,0031
$t_2$	$\theta_{21}$	-0,1725
	$\lambda_{21}$	0,1736
$t_3$	$\theta_{31}$	-0,0338
	$\lambda_{31}$	0,0338

Berdasarkan hasil estimasi pada Tabel 4.6 dan Tabel 4.7 maka model regresi semiparametrik campuran *Spline truncated* dengan 1 titik knot dan kernel dapat ditulis menjadi :

$$\begin{aligned}\hat{y}_i = & 2,2482 + 5,236x_i + 0,0034t_{1i} - 0,0031(t_{1i} - 143,285)_+ - 0,1725t_{2i} + \\ & 0,1736(t_{2i} - 18,095)_+ - 0,0338t_{3i} + 0,0338(t_{3i} - 36,429)_+ + \\ & \sum_{i=1}^{691} \frac{\frac{1}{2,948} K\left(\frac{z - z_i}{2,948}\right)}{\sum_{i=1}^n \frac{1}{2,948} K\left(\frac{z - z_i}{2,948}\right)} Y_i\end{aligned}$$

#### 4.3.2.2 Model dengan Komponen *Spline Truncated* 2 Titik Knot

Berikut ini model regresi semiparametrik campuran *Spline truncated* linear dan kernel dengan masing-masing 2 titik knot pada komponen *Spline truncated*.

$$\begin{aligned}y_i = & \beta_0 + \beta_1 x_i + \theta_{11} t_{1i} + \lambda_{11}(t_{1i} - K_{11})_+ + \lambda_{12}(t_{1i} - K_{12})_+ + \theta_{21} t_{2i} + \\ & \lambda_{21}(t_{2i} - K_{21})_+ + \lambda_{22}(t_{2i} - K_{22})_+ + \theta_{31} t_{3i} + \lambda_{31}(t_{3i} - K_{31})_+ + \\ & \lambda_{32}(t_{3i} - K_{32})_+ + \sum_{i=1}^n \frac{\frac{1}{\alpha} K\left(\frac{z - z_i}{\alpha}\right)}{\sum_{i=1}^n \frac{1}{\alpha} K\left(\frac{z - z_i}{\alpha}\right)} Y_i + \varepsilon_i\end{aligned}$$

Nilai GCV yang dihasilkan dapat dilihat pada Tabel 4.8.

Tabel 4.8 Perbandingan nilai GCV Model Komponen *Spline* dengan 2 Titik Knot

Parametrik Linear	<i>Spline</i>			Kernel	$GCV = (K_1, K_2, K_3, \alpha)$
Beta 1	Knot (K)			Bandwidth ( $\alpha$ )	
$x$	$t_1 = K_1$	$t_2 = K_1$	$t_3 = K_1$	$\alpha$	
	$t_1 = K_2$	$t_2 = K_2$	$t_3 = K_2$		
5,2803	184,1254	20	52,6923	3,03099	0,00021113277*
	612,9470	40	223,4615		
5,2816	184,1254	20	52,6923	2,9129	0,00021138408
	398,5362	30	138,0769		
5,2820	184,1254	20	52,6923	3,0403	0,00021154772
	720,1523	45	266,1539		
5,2757	184,1254	20	52,6923	3,0015	0,00021192542
	505,7415	35	180,7692		
5,2755	184,1254	20	52,6923	3,01474	0,00021247339
	827,3577	50	308,8462		



Berdasarkan Tabel 4.8 didapat GCV terkecil 0,00021113277. Berdasarkan GCV terkecil ini memberikan estimasi parameter-parameter pada komponen *Spline* disajikan pada Tabel 4.9 :

Tabel 4.9 Estimasi Parameter Komponen *Spline Truncated* dengan 2 Titik Knot

Variabel	Parameter	Estimasi
$t_1$	$\theta_{11}$	0,0025
	$\lambda_{11}$	-0,0021
	$\lambda_{12}$	-0,0003
$t_2$	$\theta_{21}$	-0,12027
	$\lambda_{21}$	0,1293
	$\lambda_{22}$	-0,0101
$t_3$	$\theta_{31}$	-0,0158
	$\lambda_{31}$	0,01645
	$\lambda_{32}$	-0,0012

Berdasarkan hasil estimasi pada Tabel 4.8 dan Tabel 4.9 maka model regresi semiparametrik campuran *Spline truncated* dengan 2 titik knot dan kernel dapat ditulis menjadi :

$$\begin{aligned} \hat{y}_i = & 0.94436929 + 5,2803x_i + 0,0025t_{1i} - 0,0021(t_{1i} - 184,1254)_+ + \\ & -0,0003(t_{1i} - 612,9470)_+ - 0,12027t_{2i} + 0,1293(t_{2i} - 20)_+ + \\ & -0,0101(t_{2i} - 40)_+ - 0,0158t_{3i} + 0,01645(t_{3i} - 52,6923)_+ + \\ & 0,0012(t_{3i} - 223,4615)_+ + \sum_{i=1}^{691} \frac{1}{3,03099} K\left(\frac{z - z_i}{3,03099}\right) \\ & \sum_{i=1}^n \frac{1}{3,03099} K\left(\frac{z - z_i}{3,03099}\right) Y_i \end{aligned}$$

#### 4.3.2.3 Model dengan Komponen *Spline Truncated* 3 Titik Knot

Berikut ini model regresi semiparametrik campuran *Spline truncated* linear dan kernel dengan masing-masing 3 titik knot pada komponen *Spline truncated* :

$$y_i = \beta_0 + \beta_1 x_i + \theta_{11} t_{1i} + \lambda_{11} (t_{1i} - K_{11})_+ + \lambda_{12} (t_{1i} - K_{12})_+ + \lambda_{13} (t_{1i} - K_{13})_+ + \\ \theta_{21} t_{2i} + \lambda_{21} (t_{2i} - K_{21})_+ + \lambda_{22} (t_{2i} - K_{22})_+ + \lambda_{23} (t_{2i} - K_{23})_+ + \theta_{31} t_{3i} + \\ \lambda_{31} (t_{3i} - K_{31})_+ + \lambda_{32} (t_{3i} - K_{32})_+ + \lambda_{33} (t_{3i} - K_{33})_+ + \sum_{i=1}^n \frac{\frac{1}{\alpha} K\left(\frac{z - z_i}{\alpha}\right)}{\sum_{i=1}^n \frac{1}{\alpha} K\left(\frac{z - z_i}{\alpha}\right)} Y_i + \varepsilon_i$$

Nilai GCV yang dihasilkan dapat dilihat pada Tabel 4.10.

Tabel 4.10 Perbandingan nilai GCV Model Komponen *Spline* dengan 3 Titik Knot

Parametrik	<i>Spline</i>			Kernel	$GCV = (K_1, K_2, K_3, \alpha)$
Linear					
Beta 1	Knot(K)			Bandwidth ( $\alpha$ )	
$x$	$t_1 = K_1$	$t_2 = K_1$	$t_3 = K_1$	$\alpha$	
	$t_1 = K_2$	$t_2 = K_2$	$t_3 = K_2$		
	$t_1 = K_3$	$t_2 = K_3$	$t_3 = K_3$		
5,2964	203,617	20,909	60,455	3,0135	0,00021036855*
	330,315	26,818	110,909		
	710,406	44,546	262,273		
5,2870	203,617	20,909	60,455	3,0165	0,00021038642
	330,315	26,818	110,909		
	583,709	38,636	211,818		
5,2963	203,617	20,909	60,4545	2,9878	0,00021070912
	330,315	26,818	110,909		
	837,104	50,455	312,727		
5,2939	203,617	20,909	60,455	2,9628	0,00021076662
	330,315	26,818	110,909		
	963,801	56,364	363,182		
5,2918273	203,617	20,909	60,455	2,9586	0,00021111858
	330,315	26,818	110,909		
	1090,498	62,273	413,636		

Berdasarkan Tabel 4.10 didapat GCV terkecil 0,00021036855. Berdasarkan GCV terkecil ini memberikan estimasi parameter-parameter pada komponen *Spline* disajikan pada Tabel 4.11 :

Tabel 4.11 Estimasi Parameter Komponen *Spline Truncated* dengan 3 Titik Knot

Variabel	Parameter	Estimasi
$t_1$	$\theta_{11}$	0,0006
	$\lambda_{11}$	0,0009
	$\lambda_{12}$	-0,0012
	$\lambda_{13}$	-0,0002
$t_2$	$\theta_{21}$	-0,1097
	$\lambda_{21}$	0,1295
	$\lambda_{22}$	-0,0148
	$\lambda_{23}$	-0,0067
$t_3$	$\theta_{31}$	-0,0153
	$\lambda_{31}$	0,0191
	$\lambda_{32}$	-0,0038
	$\lambda_{33}$	-0,0005

Berdasarkan hasil estimasi pada Tabel 4.10 dan Tabel 4.11 maka model regresi semiparametrik campuran *Spline truncated* dengan 3 titik knot dan kernel dapat ditulis menjadi :

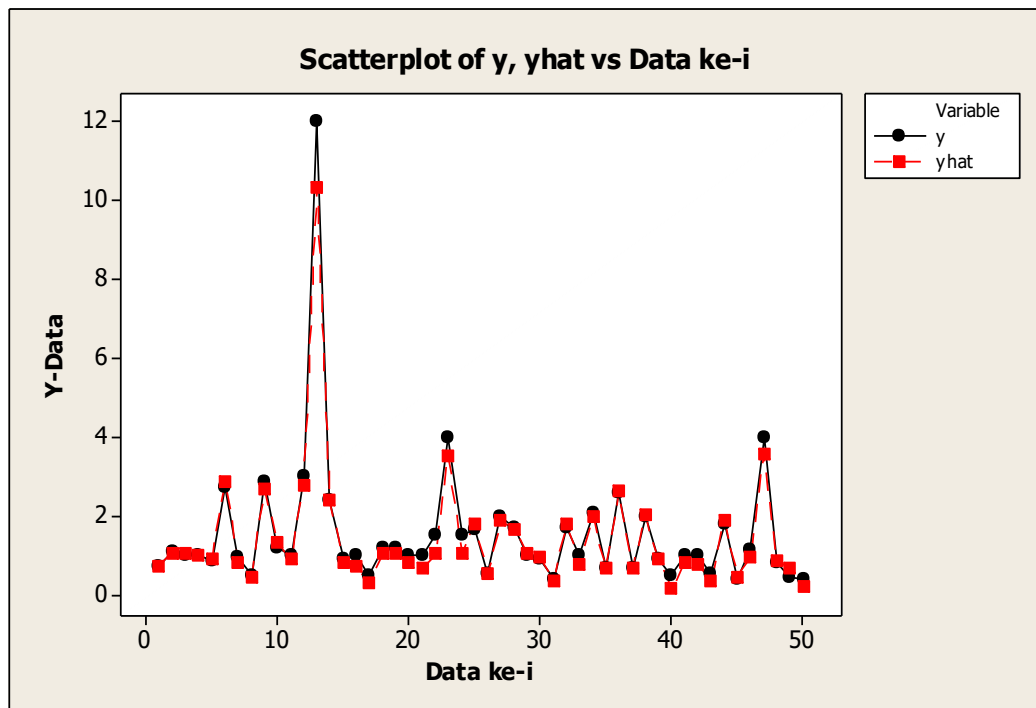
$$\begin{aligned}
\hat{y}_i = & 1,0369 + 5,2964x_i + 0,0006t_{1i} + 0,0009(t_{1i} - 203,6173)_+ + \\
& -0,0012(t_{1i} - 330,3146)_+ - 0,0002(t_{1i} - 710,4064)_+ + \\
& -0,1097t_{2i} + 0,1295(t_{2i} - 20,9091)_+ - 0,0148(t_{2i} - 26,8182)_+ + \\
& -0,0067(t_{2i} - 44,5455)_+ - 0,0153t_{3i} + 0,0191(t_{3i} - 60,4545)_+ + \\
& -0,0038(t_{3i} - 110,9091)_+ + -0,0005(t_{3i} - 262,2727)_+ + \\
& \sum_{i=1}^{691} \frac{1}{3,0135} K\left(\frac{z - z_i}{3,0135}\right) Y_i \\
& \sum_{i=1}^n \frac{1}{3,0135} K\left(\frac{z - z_i}{3,0135}\right)
\end{aligned}$$

Selanjutnya perbandingan nilai GCV dari ketiga knot dapat dilihat pada Tabel 4.12.

Tabel 4.12 Perbandingan GCV untuk Model dengan Komponen *Spline Truncated* 1,2 dan 3 Knot

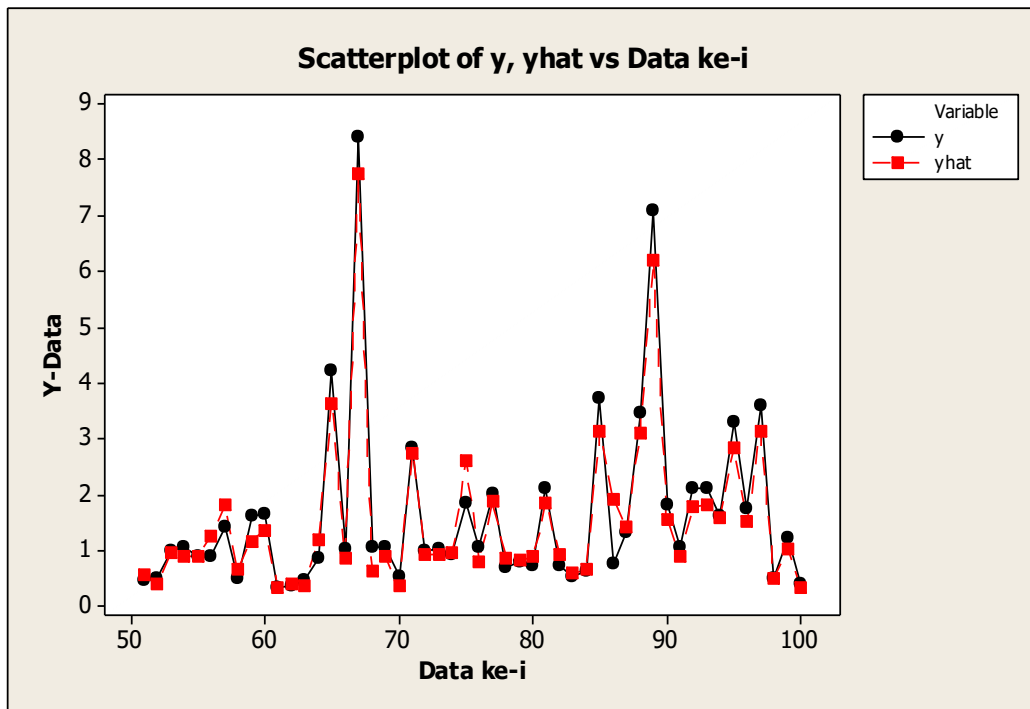
No.	Model	GCV
1	1 Titik Knot	0,00021272675
2	2 Titik Knot	0,00021113277
3	3 Titik Knot	0,00021036855*

Berdasarkan nilai GCV pada Tabel 4.6, Tabel 4.8 dan Tabel 4.10 maka diperoleh nilai GCV terkecil adalah pada model dengan komponen *Spline truncated* 3 titik knot yaitu dengan nilai GCV sebesar 0,00021036855. Dengan demikian nilai yang digunakan pada pemodelan data produksi padi di Jawa Tengah adalah model dengan 3 titik knot. Plot antara  $y$  dan  $\hat{y}$  untuk data ke-1 sampai data ke-50 dapat dilihat pada Gambar 4.6.



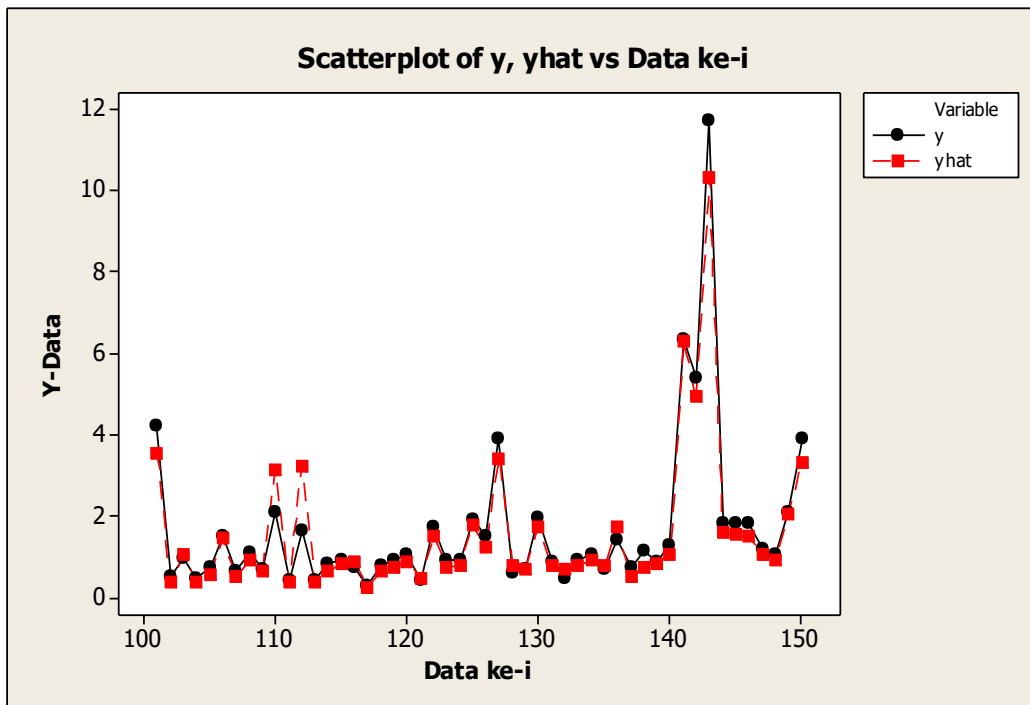
Gambar 4.6 Plot data  $y$  dan  $\hat{y}$  Data ke-1 Sampai Data ke-50

Plot antara  $y$  dan  $\hat{y}$  untuk data ke-51 sampai data ke-100 dapat dilihat pada Gambar 4.7.



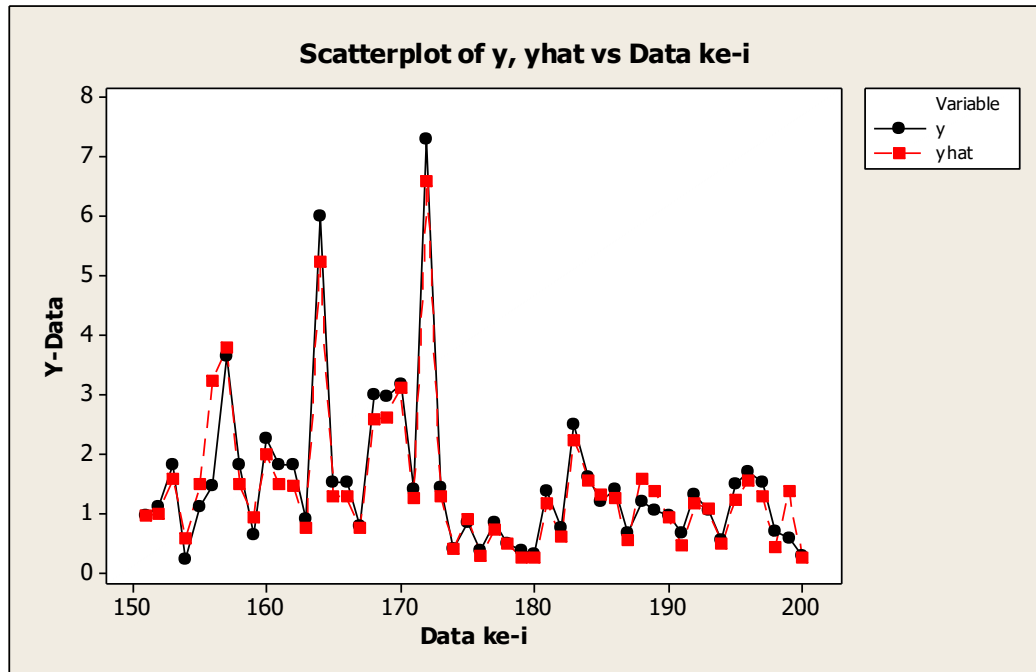
Gambar 4.7 Plot data  $y$  dan  $\hat{y}$  Data ke-51 Sampai Data ke-100

Plot antara  $y$  dan  $\hat{y}$  untuk data ke-101 sampai data ke-150 dapat dilihat pada Gambar 4.8.



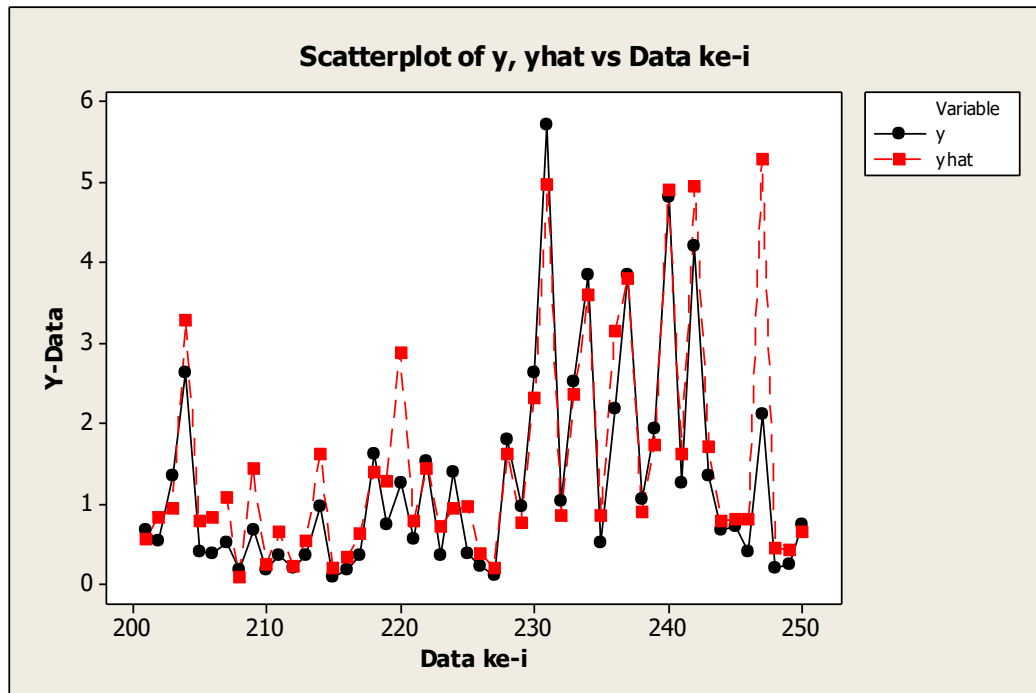
Gambar 4.8 Plot data  $y$  dan  $\hat{y}$  Data ke-101 Sampai Data ke-150

Plot antara  $y$  dan  $\hat{y}$  untuk data ke-151 sampai data ke-200 dapat dilihat pada Gambar 4.9.



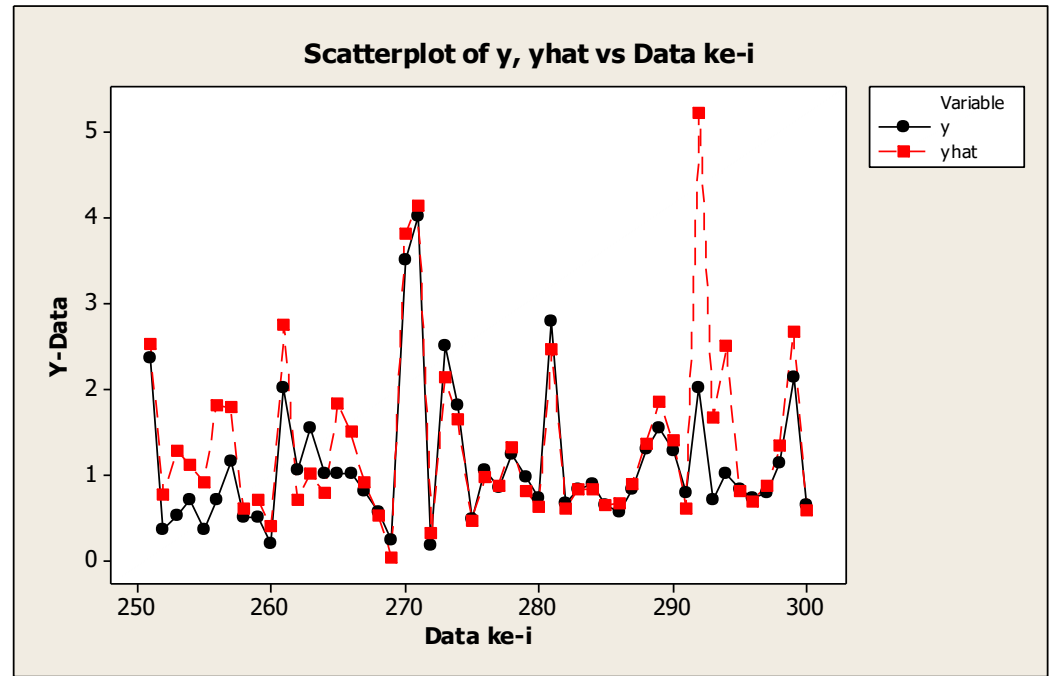
Gambar 4.9 Plot data  $y$  dan  $\hat{y}$  Data ke-151 Sampai Data ke-200

Plot antara  $y$  dan  $\hat{y}$  untuk data ke-201 sampai data ke-250 dapat dilihat pada Gambar 4.10.



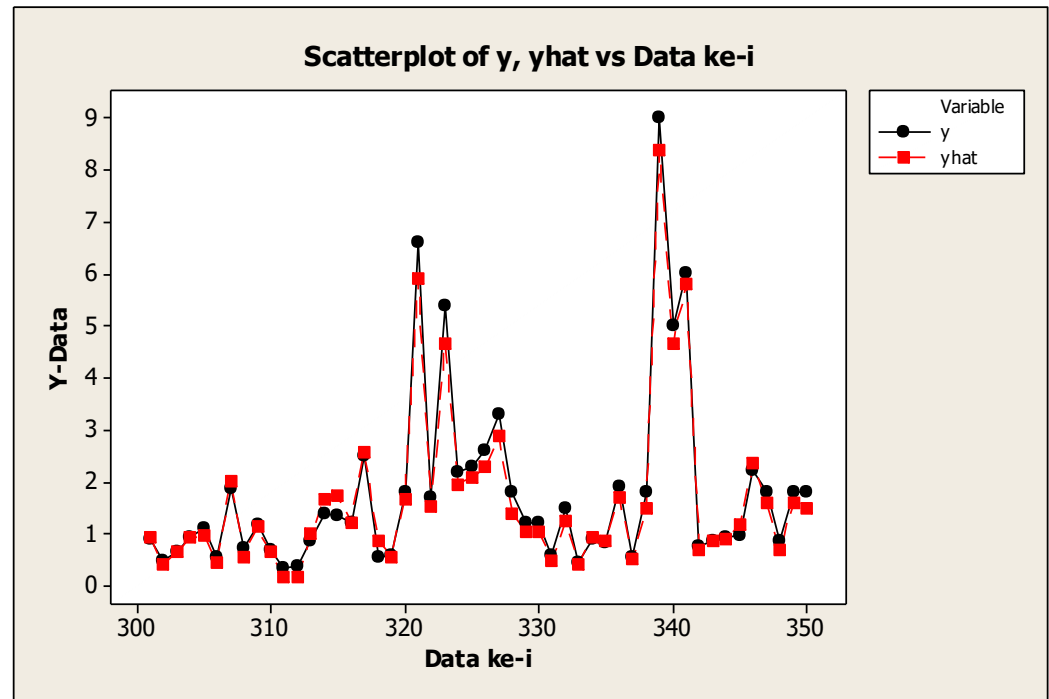
Gambar 4.10 Plot data  $y$  dan  $\hat{y}$  Data ke-201 Sampai Data ke-250

Plot antara  $y$  dan  $\hat{y}$  untuk data ke-251 sampai data ke-300 dapat dilihat pada Gambar 4.11.



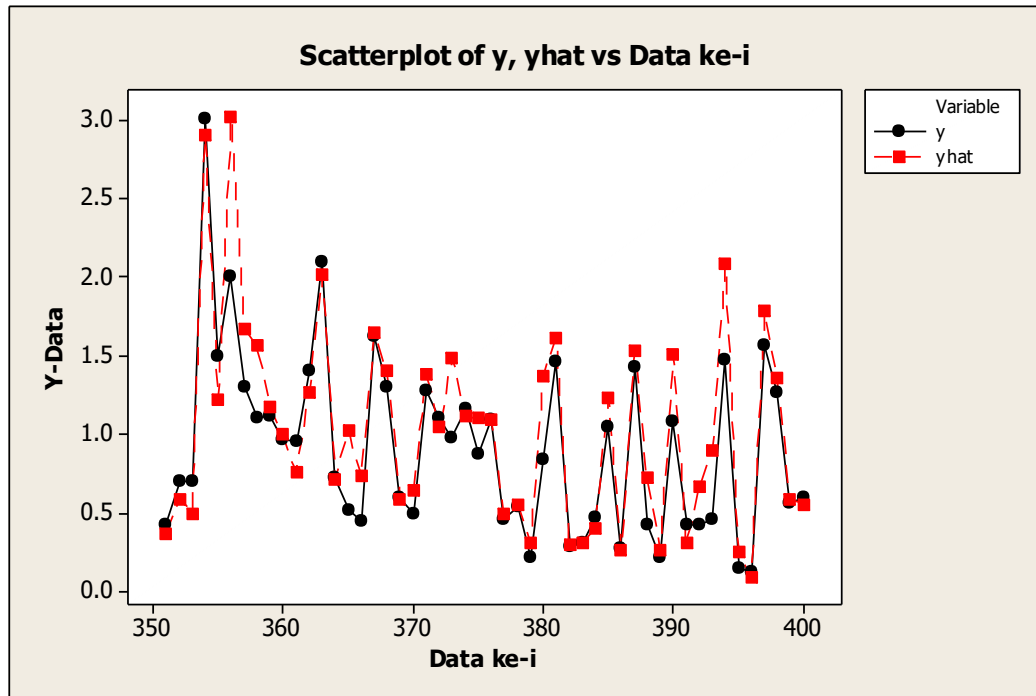
Gambar 4.11 Plot data  $y$  dan  $\hat{y}$  Data ke-251 Sampai Data ke-300

Plot antara  $y$  dan  $\hat{y}$  untuk data ke-301 sampai data ke-350 dapat dilihat pada Gambar 4.12.



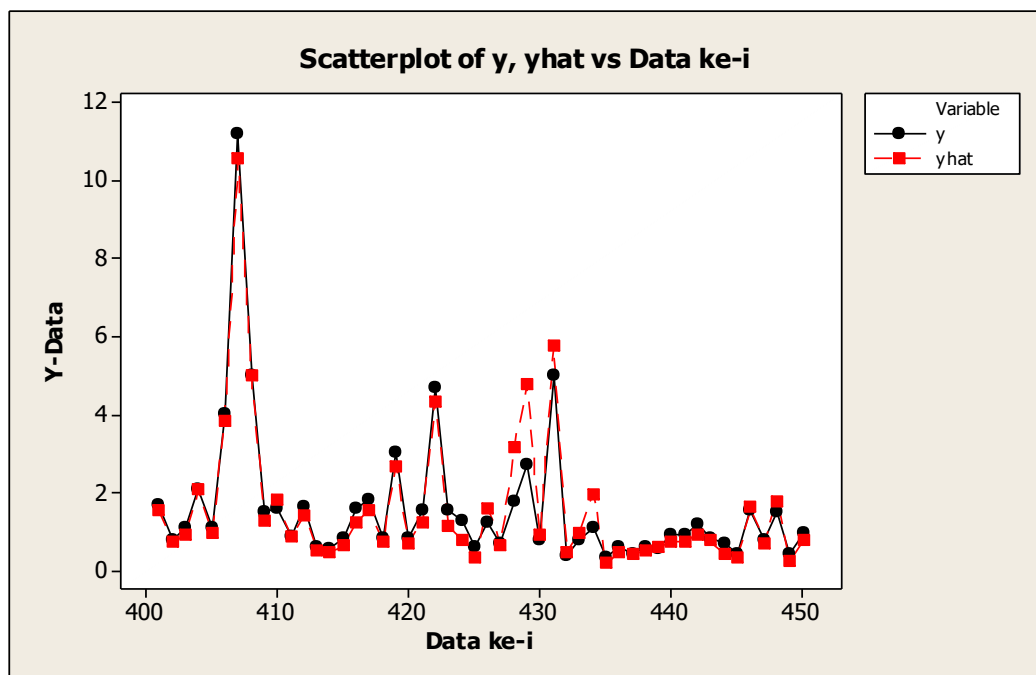
Gambar 4.12 Plot data  $y$  dan  $\hat{y}$  Data ke-301 Sampai Data ke-350

Plot antara  $y$  dan  $\hat{y}$  untuk data ke-351 sampai data ke-400 dapat dilihat pada Gambar 4.13.



Gambar 4.13 Plot data  $y$  dan  $\hat{y}$  Data ke-351 Sampai Data ke-400

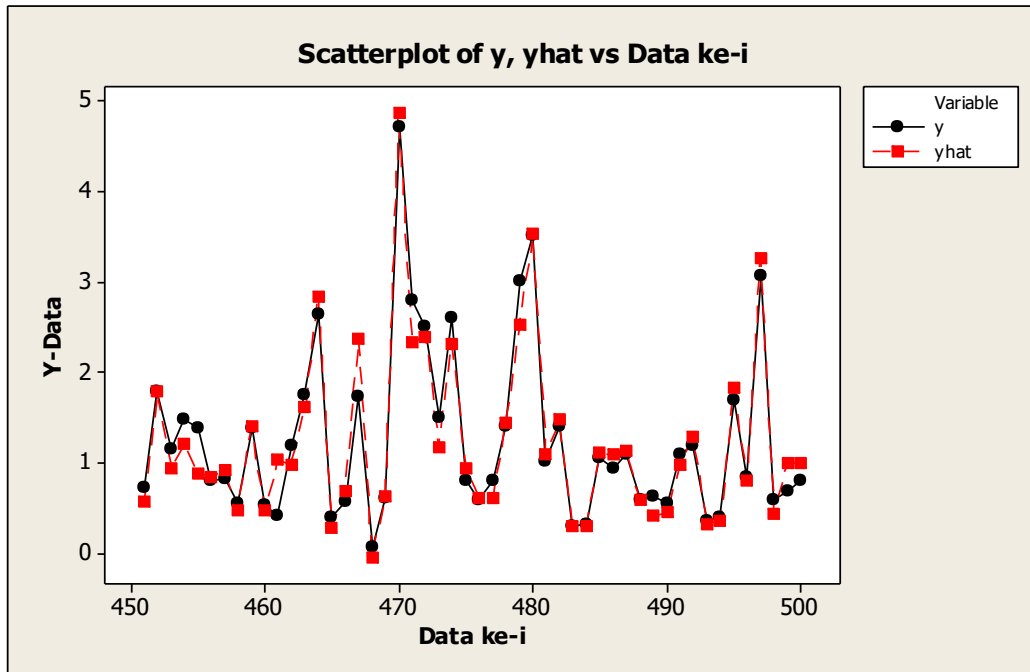
Plot antara  $y$  dan  $\hat{y}$  untuk data ke-401 sampai data ke-450 dapat dilihat pada Gambar 4.14.



Gambar 4.14 Plot data  $y$  dan  $\hat{y}$  Data ke-401 Sampai Data ke-450

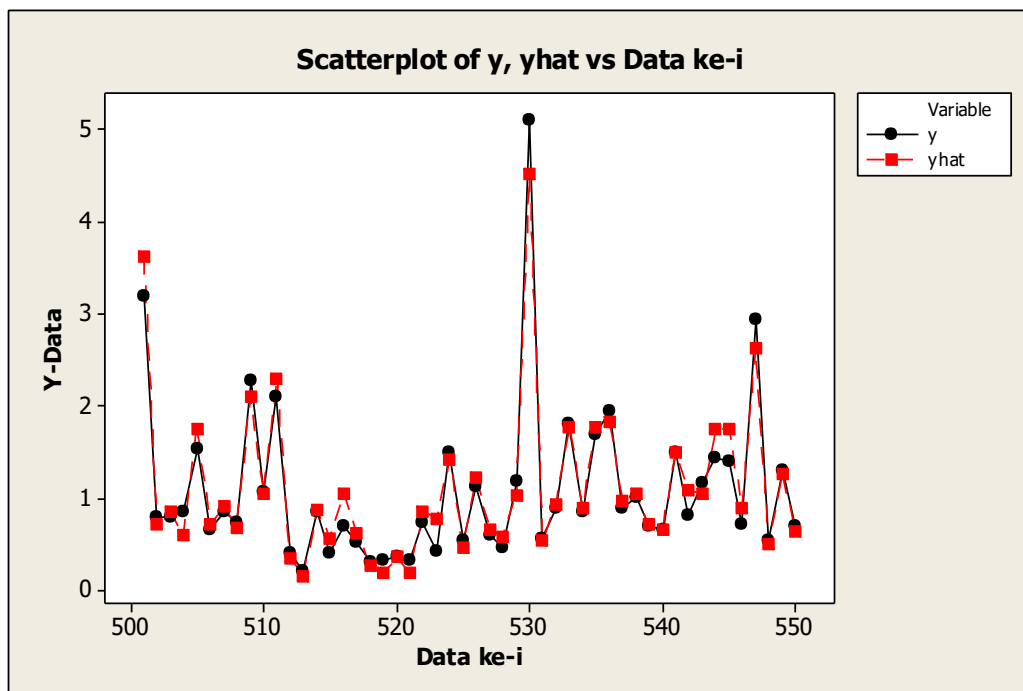


Plot antara  $y$  dan  $\hat{y}$  untuk data ke-451 sampai data ke-500 dapat dilihat pada Gambar 4.15.



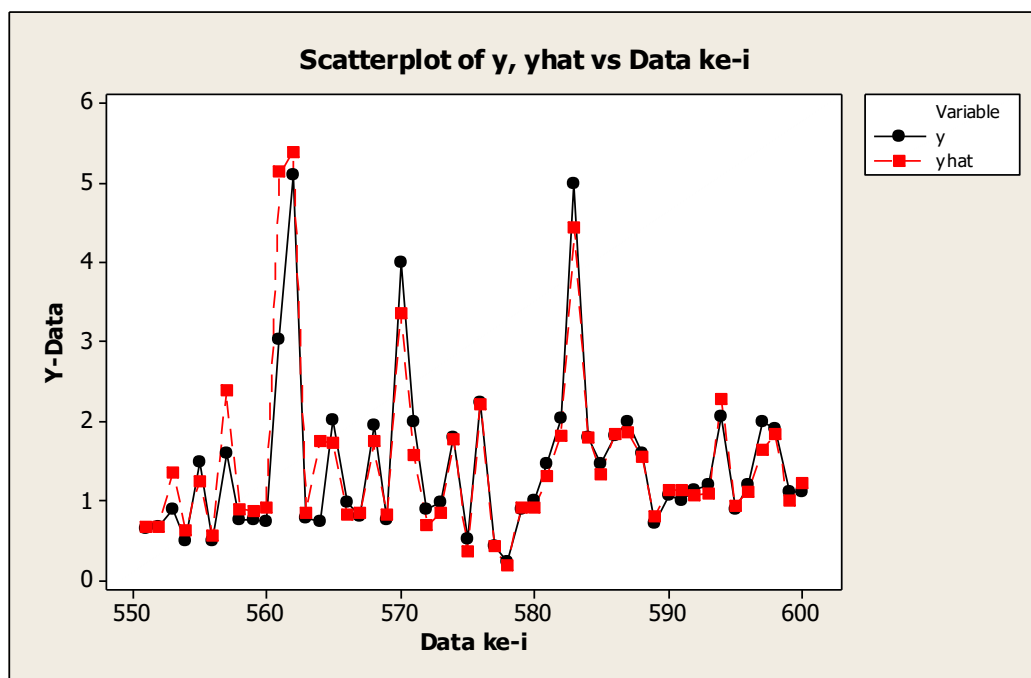
Gambar 4.15 Plot data  $y$  dan  $\hat{y}$  Data ke-451 Sampai Data ke-500

Plot antara  $y$  dan  $\hat{y}$  untuk data ke-501 sampai data ke-550 dapat dilihat pada Gambar 4.16.



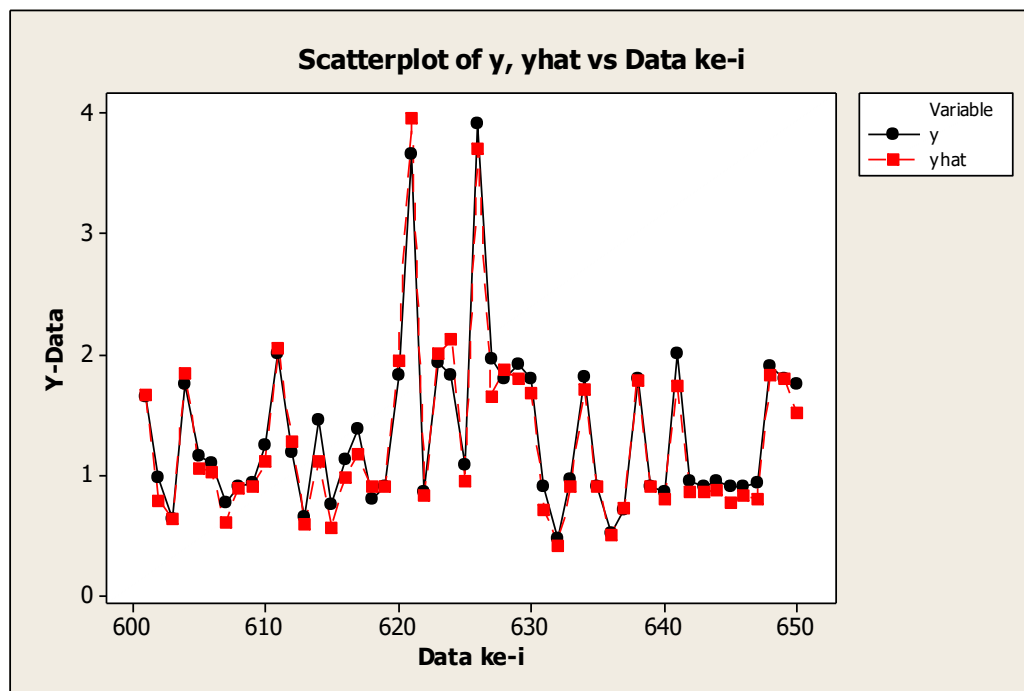
Gambar 4.16 Plot data  $y$  dan  $\hat{y}$  Data ke-501 Sampai Data ke-550

Plot antara  $y$  dan  $\hat{y}$  untuk data ke-551 sampai data ke-600 dapat dilihat pada Gambar 4.17.



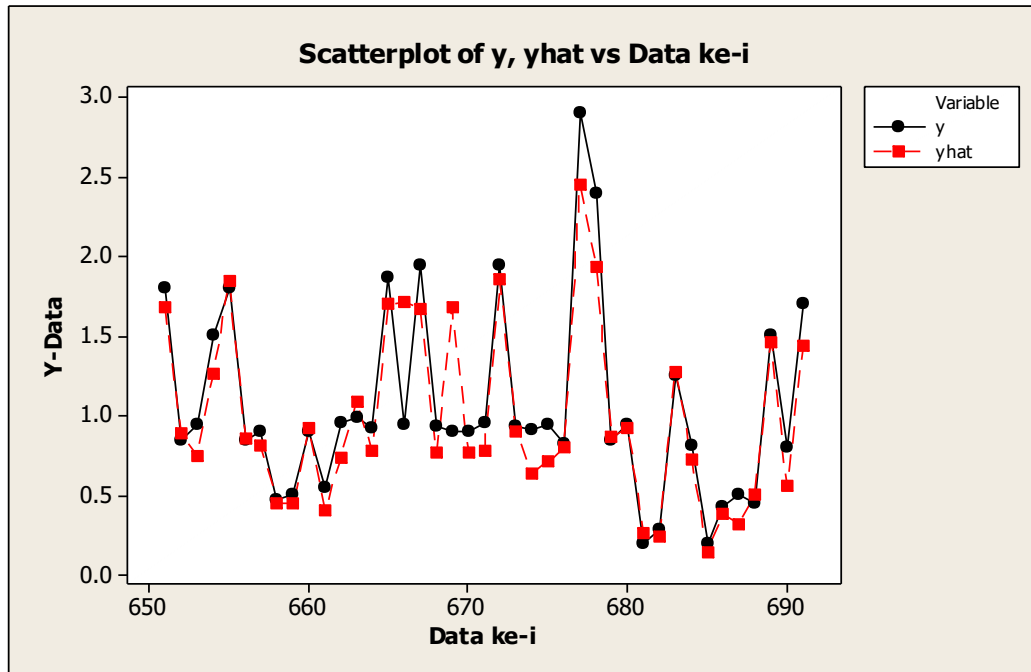
Gambar 4.17 Plot data  $y$  dan  $\hat{y}$  Data ke-551 Sampai Data ke-600

Plot antara  $y$  dan  $\hat{y}$  untuk data ke-601 sampai data ke-650 dapat dilihat pada Gambar 4.18.



Gambar 4.18 Plot data  $y$  dan  $\hat{y}$  Data ke-601 Sampai Data ke-650

Plot antara  $y$  dan  $\hat{y}$  untuk data ke-650 sampai data ke-691 dapat dilihat pada Gambar 4.19.



Gambar 4.19 Plot data  $y$  dan  $\hat{y}$  Data ke-651 Sampai Data ke-691

Berdasarkan Gambar 4.6 sampai Gambar 4.19 terlihat bahwa grafik  $\hat{y}$  mendekati grafik data sebenarnya. Dari perhitungan, model dengan 3 titik knot ini mempunyai nilai  $R^2$  sebesar 91,91. Nilai  $R^2$  tersebut lebih tinggi dibandingkan menggunakan fungsi *Spline* saja maupun Kernel saja pada komponen nonparametriknya. Tabel perbandingan nilai  $R^2$  dapat dilihat pada Tabel 4.13.

Tabel 4.13 Perbandingan Nilai  $R^2$

Fungsi	$R^2$
<i>Spline Truncated</i>	91.89 %
Kernel	91.47 %
Campuran <i>Spline Truncated</i> dan Kernel	91,91 %

Terlihat bahwa nilai  $R^2$  model dengan komponen nonparametrik menggunakan fungsi atau estimator campuran *Spline Truncated* dan Kernel sebesar 91,91 % yang berarti variabel luas panen, penggunaan pupuk, penggunaan benih, penggunaan tenaga kerja dan penggunaan pestisida mampu menjelaskan variabel respon produksi padi sebesar 91,91 %.

#### 4.3.2 Interpretasi Model Regresi Semiparametrik Campuran

Model untuk pola data yang mengikuti kurva parametrik linear dan *Spline truncated* dapat diinterpretasikan. Sementara itu, untuk pola data yang mengikuti kernel tidak dapat diinterpretasikan. Berikut ini adalah model masing-masing kelompok data dan interpretasinya :

##### 1) Model untuk variabel luas panen

Dengan mengasumsikan data selain luas panen konstan, maka diperoleh model umumnya :

$$\hat{y}_i = 5,2964x_i + c_1$$

dimana :

$$\begin{aligned} c_1 = & 1,0369 + 0,0006t_{1i} + 0,0009(t_{1i} - 203,6173)_+ - 0,0012(t_{1i} - 330,3146)_+ + \\ & - 0,0002(t_{1i} - 710,4064)_+ - 0,1097t_{2i} + 0,1295(t_{2i} - 20,9091)_+ + \\ & - 0,0148(t_{2i} - 26,8182)_+ - 0,0067(t_{2i} - 44,5455)_+ - 0,0153t_{3i} + \\ & 0,0191(t_{3i} - 60,4545)_+ - 0,0038(t_{3i} - 110,9091)_+ - 0,0005(t_{3i} - 262,2727)_+ + \\ & \frac{\sum_{i=1}^{691} \frac{1}{3,0135} K\left(\frac{z - z_i}{3,0135}\right)}{\sum_{i=1}^n \frac{1}{3,0135} K\left(\frac{z - z_i}{3,0135}\right)} Y_i \end{aligned}$$

Sehingga dengan mengasumsikan data selain luas panen konstan disimpulkan bahwa ketika luas panen naik sebesar 1 hektar maka produksi padi akan naik sebesar 5,2964 ton.

##### 2) Model untuk variabel penggunaan pupuk

Dengan mengasumsikan data selain penggunaan pupuk konstan, maka diperoleh model umumnya :

$$\begin{aligned} \hat{y}_i = & 0,0006t_{1i} + 0,0009(t_{1i} - 203,6173)_+ - 0,0012(t_{1i} - 330,3146)_+ + \\ & - 0,0002(t_{1i} - 710,4064)_+ + c_2 \end{aligned}$$

dimana :

$$\begin{aligned}
 c_2 = & 1,0369 + 5,2964x - 0,1097t_{2i} + 0,1295(t_{2i} - 20,9091)_+ + \\
 & - 0,0148(t_{2i} - 26,8182)_+ - 0,0067(t_{2i} - 44,5455)_+ + \\
 & - 0,0153t_{3i} + 0,0191(t_{3i} - 60,4545)_+ - 0,0038(t_{3i} - 110,9091)_+ + \\
 & - 0,0005(t_{3i} - 262,2727)_+ + \sum_{i=1}^{691} \frac{\frac{1}{3,0135} K\left(\frac{z - z_i}{3,0135}\right)}{\sum_{i=1}^n \frac{1}{3,0135} K\left(\frac{z - z_i}{3,0135}\right)} Y_i
 \end{aligned}$$

Untuk interpretasi model ini digunakan fungsi seperti berikut :

$$\hat{y} = \begin{cases} 0,0006t_{li} + c_2, & t_{li} \leq 203,6173 \\ -0,18326 + 0,0015t_{li} + c_2, & 203,6173 < t_{li} \leq 330,3146 \\ 0,213122 + 0,0003t_{li} + c_2, & 330,3146 < t_{li} \leq 710,4064 \\ 0,355203 + 0,0001t_{li} + c_2, & t_{li} > 710,4064 \end{cases}$$

- Untuk penggunaan pupuk kurang dari 203,6173 kg, ketika dilakukan penambahan pupuk sebesar 1 kg maka produksi padi naik sebesar 0,0006 ton.
- Untuk penggunaan pupuk berada di rentang nilai 203,6173 sampai 330,3146 kg, ketika dilakukan penambahan pupuk sebesar 1 kg maka produksi padi naik sebesar 0,0015 ton.
- Untuk penggunaan pupuk berada di rentang nilai 330,3146 sampai 710,4064 kg, ketika dilakukan penambahan pupuk sebesar 1 kg maka produksi padi naik sebesar 0,0001 ton.
- Untuk penggunaan pupuk lebih dari dari 710,4064 kg, ketika dilakukan penambahan pupuk sebesar 1 kg maka produksi padi naik sebesar 0,0001 ton.

### 3) Model untuk variabel penggunaan benih

Dengan mengasumsikan data selain penggunaan benih konstan, maka diperoleh model umumnya :

$$\begin{aligned}
 \hat{y}_i = & -0,1097t_{2i} + 0,1295(t_{2i} - 20,9091)_+ - 0,0148(t_{2i} - 26,8182)_+ + \\
 & - 0,0067(t_{2i} - 44,5455)_+ + c_3
 \end{aligned}$$

dimana :

$$\begin{aligned}
 c_3 = & 1,0369 + 5,2964x_i + 0,0006t_{1i} + 0,0009(t_{1i} - 203,6173)_+ + \\
 & -0,0012(t_{1i} - 330,3146)_+ - 0,0002(t_{1i} - 710,4064)_+ + \\
 & -0,0153t_{3i} + 0,0191(t_{3i} - 60,4545)_+ - 0,0038(t_{3i} - 110,9091)_+ + \\
 & -0,0005(t_{3i} - 262,2727)_+ + \sum_{i=1}^{691} \frac{\frac{1}{3,0135} K\left(\frac{z - z_i}{3,0135}\right)}{\sum_{i=1}^n \frac{1}{3,0135} K\left(\frac{z - z_i}{3,0135}\right)} Y_i
 \end{aligned}$$

Untuk interpretasi model ini digunakan fungsi seperti berikut :

$$\hat{y} = \begin{cases} -0,1097t_{2i} + c_3, & t_{2i} \leq 20,9091 \\ -2,70773 + 0,01985t_{2i} + c_3, & 20,9091 < t_{2i} \leq 26,8182 \\ -2,31082 + 0,005t_{2i} + c_3, & 26,8182 < t_{2i} \leq 44,5455 \\ -2,01236 - 0,0017t_{2i} + c_3, & t_{2i} > 44,5455 \end{cases}$$

- Untuk penggunaan benih kurang dari 20,9091 kg, dilakukan penambahan benih sebesar 1 kg maka produksi padi turun sebesar 0,1097 ton.
- Untuk penggunaan benih berada di rentang nilai 20,9091 sampai 26,8182 kg , ketika dilakukan penambahan benih sebesar 1 kg maka produksi padi naik sebesar 0,01985 ton.
- Untuk penggunaan benih berada di rentang nilai 26,8182 sampai 44,5455 kg, ketika penggunaan benih naik sebesar 1 kg maka produksi padi naik sebesar 0,005 ton.
- Untuk penggunaan benih lebih dari dari 44,5455 kg, ketika penggunaan benih naik sebesar 1 kg maka produksi padi turun sebesar 0,001 ton.

#### 4) Model untuk variabel penggunaan tenaga kerja

Dengan mengasumsikan data selain penggunaan tenaga kerja konstan, maka diperoleh model umumnya :

$$\begin{aligned}
 \hat{y}_i = & -0,0153t_{3i} + 0,0191(t_{3i} - 60,4545)_+ - 0,0038(t_{3i} - 110,9091)_+ + \\
 & -0,0005(t_{3i} - 262,2727)_+ + c_4
 \end{aligned}$$

dimana :

$$\begin{aligned}
c_4 = & 1,0369 + 5,2964x_i + 0,0006t_{1i} + 0,0009(t_{1i} - 203,6173)_+ + \\
& -0,0012(t_{1i} - 330,3146)_+ - 0,0002(t_{1i} - 710,4064)_+ + \\
& -0,0153t_{3i} + 0,0191(t_{3i} - 60,4545)_+ - 0,0038(t_{3i} - 110,9091)_+ + \\
& -0,0005(t_{3i} - 262,2727)_+ + \sum_{i=1}^{691} \frac{\frac{1}{3,0135} K\left(\frac{z - z_i}{3,0135}\right)}{\sum_{i=1}^n \frac{1}{3,0135} K\left(\frac{z - z_i}{3,0135}\right)} y_i
\end{aligned}$$

Untuk interpretasi model ini digunakan fungsi seperti berikut :

$$\hat{y} = \begin{cases} -0,0153t_{3i} + c_4, & t_{2i} \leq 60,4545 \\ -1,15468 + 0,0038t_{3i} + c_4, & 60,4545 < t_{2i} \leq 110,9091 \\ -0,73323 + 0,0001t_{3i} + c_4, & 110,9091 < t_{2i} \leq 262,2727 \\ -0,60209 - 0,0005t_{3i} + c_4, & t_{2i} > 262,2727 \end{cases}$$

- Untuk penggunaan tenaga kerja kurang dari 60 orang, ketika penggunaan tenaga kerja bertambah sebanyak 1 orang maka produksi padi turun sebesar 0,0153 ton.
- Untuk penggunaan tenaga kerja berada di rentang nilai 60 sampai 110 orang, ketika penggunaan tenaga kerja bertambah sebanyak 1 orang maka produksi padi naik sebesar 0,0038 ton.
- Untuk penggunaan tenaga kerja berada di rentang nilai 110 sampai 262 orang, ketika penggunaan tenaga kerja bertambah sebanyak 1 orang maka produksi padi naik sebesar 0,0001 ton.
- Untuk penggunaan tenaga kerja lebih dari 262 orang, ketika penggunaan tenaga kerja bertambah sebanyak 1 orang maka produksi padi turun sebesar 0,0005 ton.

Pemodelan data menggunakan estimator campuran *Spline truncated* dan kernel pada data produksi padi tersebut menghasilkan nilai  $R^2$  sebesar 91,91%. Dari hasil diketahui bahwa luas panen berhubungan positif terhadap produksi padi, dimana semakin luas lahan panen maka semakin banyak pula produksi padi yang dihasilkan. Sementara itu untuk penggunaan pupuk, penggunaan benih dan

penggunaan tenaga kerja perlu digunakan secara optimal sehingga dapat menghasilkan produksi padi yang optimal pula.



## BAB V

### KESIMPULAN DAN SARAN

#### 1.1 Kesimpulan

1. Diberikan data berpasangan  $(x_i, t_i, z_i, y_i)$ ,  $i = 1, 2, \dots, n$ , yang diasumsikan mengikuti model regresi semiparametrik  $y_i = \mu(x_i, t_i, z_i) + \varepsilon_i$ . Kurva regresi  $\mu$  diasumsikan bersifat additif, sehingga dapat ditulis menjadi  $\mu(x_i, t_i, z_i) = f(x_i) + g(t_i) + h(z_i)$ . Kurva regresi  $f(x_i)$  dihampiri dengan fungsi parametrik linear,  $g(t_i)$  dihampiri dengan fungsi *Spline truncated* linear dan kurva regresi  $h(z_i)$  dapat dihampiri dengan fungsi kernel. Model regresi semiparametrik campuran dengan komponen nonparametrik multivariabel diberikan oleh :

$$y_i = f(x_i) + \sum_{r=1}^p g_r(t_{ri}) + \sum_{s=1}^q h_s(z_{si})$$

dalam bentuk matriks dapat ditulis :

$$\tilde{y} = \mathbf{X}\tilde{\beta} + \mathbf{G}(\tilde{k})\tilde{\theta} + \mathbf{D}(\tilde{\alpha})\tilde{y}$$

Estimator regresi semiparametrik campuran *Spline truncated* dan kernel diberikan oleh :

$$\begin{aligned}\hat{\tilde{y}} &= \hat{\tilde{\mu}}(x, t, z) \\ &= \mathbf{X}\hat{\tilde{\beta}} + \mathbf{G}(\tilde{k})\hat{\tilde{\theta}} + \mathbf{D}(\tilde{\alpha})\tilde{y}\end{aligned}$$

dengan  $\hat{\tilde{\beta}} = \mathbf{A}(\tilde{k}, \tilde{\alpha})\tilde{y}$ ,  $\hat{\tilde{\theta}} = \mathbf{B}(\tilde{k}, \tilde{\alpha})\tilde{y}$ ,  $\hat{f}(x) = \mathbf{C}(\tilde{k}, \tilde{\alpha})\tilde{y}$ ,  $\hat{g}(t) = \mathbf{K}(\tilde{k}, \tilde{\alpha})\tilde{y}$  dan

$\hat{h}(z) = \mathbf{D}(\tilde{\alpha})\tilde{y}$ . Selanjutnya estimator kurva campuran diperoleh :

$$\hat{\tilde{\mu}}(x, t, z) = \mathbf{M}(\tilde{k}, \tilde{\alpha})\tilde{y},$$

dengan  $\mathbf{M}(\tilde{k}, \tilde{\alpha}) = \mathbf{C}(\tilde{k}, \tilde{\alpha}) + \mathbf{K}(\tilde{k}, \tilde{\alpha}) + \mathbf{D}(\tilde{\alpha})$ .

2. Model regresi semiparametrik campuran *Spline truncated* dan kernel diterapkan pada data produksi padi di Jawa Tengah dengan variabel responnya adalah produksi padi ( $y$ ), sementara itu variabel prediktornya adalah luas panen ( $x$ ) yang dihampiri dengan parametrik linear,

penggunaan pupuk ( $t_1$ ), penggunaan benih ( $t_2$ ) dan penggunaan tenaga kerja ( $t_3$ ) dihampiri dengan fungsi *Spline truncated* linear, serta penggunaan pestisida ( $z$ ) dihampiri dengan fungsi kernel. Model terbaik diperoleh dari model yang memiliki nilai GCV paling kecil di antara model dengan komponen *Spline truncated* 1 titik knot, 2 titik knot dan 3 titik knot, yaitu model dengan komponen *Spline truncated* 3 titik knot. Pemodelan data menggunakan estimator campuran *Spline truncated* dan kernel pada data produksi padi tersebut menghasilkan nilai  $R^2$  sebesar 91,91%. Dari hasil diketahui bahwa luas panen berhubungan positif terhadap produksi padi, dimana semakin luas lahan panen maka semakin banyak pula produksi padi yang dihasilkan. Sementara itu untuk penggunaan pupuk, penggunaan benih dan penggunaan tenaga kerja perlu digunakan secara optimal sehingga dapat menghasilkan produksi padi yang optimal pula.

## 5.2 Saran

Penelitian ini menggunakan campuran dua kurva regresi nonparametrik, sehingga untuk penelitian berikutnya dengan data yang berbeda dapat menggunakan campuran lebih dari dua komponen nonparametrik yang sesuai dengan pola data. Selain itu, dapat pula dikembangkan dengan menggunakan fungsi *Spline truncated* kuadratik maupun kubik dan untuk kernel dapat menggunakan kernel selain Gaussian misalnya kernel *uniform*, *triangle* dan lain sebagainya.

## DAFTAR PUSTAKA

- Adyana, I. G. 2011. Estimator *Spline* dalam Regresi Nonparametrik Multirespon. *Theses* : Institut Teknologi Sepuluh November Surabaya.
- Amelia, D. 2014. Penerapan Regresi Semiparametrik *Spline* pada Pemodelan Nilai UN Siswa SMKN 1 Nguling Pasuruan. *Theses*: Institut Teknologi Sepuluh November Surabaya.
- Asmin, S. 2009. Estimator *Spline* pada Model Semiparametrik. *Theses*: Institut Teknologi Sepuluh November Surabaya.
- Asriani, L.J. 2013. Model Regresi Semiparametrik Deret Fourier. *Theses*: Institut Teknologi Sepuluh November Surabaya.
- Budiantara, I.N. 2000. Metode U, GML, CV dan GCV Dalam Regresi Nonparametrik Spline. *Majalah Ilmiah Himpunan Matematika Indonesia (MIHMI)*, 6: 285-290.
- Budiantara, I.N. 2006. Model *Spline* dengan Knots Optimal. *Jurnal Ilmu Dasar FMIPA Universitas Jember*, 7, 77-85.
- Budiantara, I N. dan Mulianah. 2007, Pemilihan Bandwidth Optimal dalam Regresi Semiparametrik Kernel dan Aplikasinya, *Sigma*, Vol. 10, No. 2, Hal. 159-166.
- Budiantara, I.N. 2009. *Spline* dalam Regresi Nonparametrik dan Semiparametrik, Sebuah Pemodelan Statistika Masa Kini dan Masa mendatang, Pidato Pengukuhan Guru Besar, Institut Teknologi Sepuluh Nopember, Surabaya.
- Budiantara, I.N., Ratnasari, V., Ratna, M., and Zain, I. 2015. The Combination of Spline and Kernel estimator for Nonparametrik Regression and Its Properties, *Applied Mathematical Science*, 9, No 122, 6083-6094.
- Chamidah, N. 2008. Inferensi Kurva Regresi Nonparametrik Berdasarkan Estimator Polinomial Lokal dengan *Error Lognormal*, *Jurnal Penelitian Eksakta*, Vol. 7, No. 1, hlm 61-69.
- Cox, D. & O'Sullivan, F. 1996. *Penalized Type Estimator for Generalized Nonparametric Regression*, 1983. *Journal of Multivariate Analysis*, 4(56): 185-206.
- Draper, N.R. & Smith, H. 1992. *Applied Regression Analysis, Second Edition*. John Wiley and sons, Inc. New York.
- Eichner, G., & Stute, W. 2012. Kernel Adjusted Nonparametric regression. *Journal of Statistical Planning and Inference*, No. 142, hal 2537-2544.
- Eubank, R.L. 1988. *Spline Smoothing and Nonparametric Regression*. Marcel Dekker : New York.
- Eubank, R. L. 1999. *Spline Smoothing and Nonparametric Regression*. Marcel Dekker: New York.
- Hardle, W. 1990. *Applied Nonparametric Regression*. Cambridge University Press: New York.
- Hazelton, M. L., & Turlach, B. A. 2011. Semiparametric Regression with Shape-constrained Penalized Splines. *Computational Statistics and Data Analysis*, No. 55, hal 2871-2879.

- Kulasekera, Q. Z., & Gallagher, C. 2011. Local Adaptive Smoothing in Kernel Regression Estimation. *Statistics and Probability Letters*, No. 88, hal 540-547.
- Muller, H. G. 1984. Optimal Designs for Nonparametric Kernel Regression. *Statistics and Probability Letters*, No. 2, hal 285-290.
- Nafi, M. 2010. Estimasi Interval Spline dalam regresi Nonparametrik. *Theses*: Institut Teknologi Sepuluh November Surabaya.
- Okumura, H., & Naito, K. 2006. Non-parametric Kernel Regression for Multinomial Data. *Journal of Multivariate Analysis*, No. 97, hal 2009-2022.
- Purnomo, A. A. 2016. Estimator Campuran Kernel dan Regresi Spline Truncated Linear Multivariabel dalam Regresi Nonparametrik. *Theses* : Institut Teknologi Sepuluh November Surabaya.
- Ratnasari, V. Budiantara, I. N., Zain, I. Ratna, M. & Mariati, N. P .A . M. 2015. Comparison Truncated Spline and Fourier Series in Multivariable Nonparametric Regression Model (Application: Data of Poverty in Papua, Indonesia). *International Journal of Basic & Applied Sciences IJBAS-IJENS* Vol : 15 N0:04, 9-14.
- Rory. 2016. Regresi Campuran Nonparametrik Spline Linear Truncated dan Fungsi Kernel untuk Pemodelan Data Kemiskinan di Provinsi Papua. *Theses* : Institut Teknologi Sepuluh November Surabaya.
- Samsodin, M. 2011. Regresi *Spline Polinomial Truncated* Multirespon untuk Pemodelan Indikator Kemiskinan di Provinsi Jawa Timur. *Theses*: Institut Teknologi Sepuluh November Surabaya.
- Srinadi. 2002. Estimator Spline pada Model Semiparametrik. *Theses*: Institut Teknologi Sepuluh November Surabaya.
- Sudiarsa, I. W., Budiantara, I. N., Suhartono, Purnami, S. W. 2015. *Combined Estimator Fourier Series and Spline Truncated in Multivariabel Nonparametrik Regression, Applied Mathematical Science*, Vol.9, 2015, no. 100, 4997-5010, HIKARI Ltd.
- Susianto, Y. 2008. Model Regresi Semiparametrik kernel untuk Menduga Produksi Padi Sawah di Jawa tengah. *Theses*: Institut Teknologi Sepuluh November Surabaya.
- Triyanto, J. 2006. Analisis Produksi Padi di Jawa Tengah. *Theses*: Universitas Diponegoro.
- Wahba G. 1990. *Spline Models for Observational Data*, SIAM Pennsylvania.
- Wahyuningdyawati, F., Kasijadi., & Heriyanto . 2003. Tingkat Adopsi Teknologi Usaha Tani padi Lahan Sawah di Jawa Timur, *Jurnal Pengkajian dan Pengembangan Teknologi Pertanian*, Vol. 6, No. 1, hlm 40-49.
- Wand, M. P., & Jones, M. C. 1995. *Kernel Smoothing*. London: Chapman & Hall.
- Wang, R. 1985. The Dimention and Basis of Space of Multivariate Splines. *Journal of Computational and Applied Mathematics*, No. 12&13, hal 163-177.
- Weinert, H. L. 2009. A Fast Compact Algorithm for Cubic Spline Smoothing. *Computational Statistics and Data Analysis*, No. 53, hal 932-940.
- William, R. 1989. On Nonparaetric Regression with Higher-Order Kernels. *Journal of Statistical Planning and Inference*, No. 23, hal 141-151.

- Zulfikar. 2005. Komponen Bias, Variansi dan MSE Estimator Kernel dalam Regresi Nonparametrik. *Theses*: Institut Teknologi Sepuluh November Surabaya.
- Zulkarnain, I. 2004. Analisis Produksi dan Keuntungan pada Usaha Tani Padi Sawah Jawa Tengan. *Theses*: Universitas Diponegoro.
- Zuriani. 2013. Analisis Produksi Padi di Kabupaten Aceh Utara. *Theses*: Universitas Sumatera Utara.

*Halaman ini sengaja dikosongkan*

## LAMPIRAN

### Lampiran 1. Data yang digunakan

No	y	x	$t_1$	$t_2$	$t_3$	z
1	0,72	0,14	571,43	50,00	236	0,57
2	1,1	0,21	547,62	57,14	210	0,38
3	1	0,21	543,22	56,68	179	0,38
4	1	0,19	651,38	52,11	266	0,42
5	0,865	0,18	564,02	50,76	226	0,45
6	2,748	0,53	1047,62	47,62	211	0,57
7	0,965	0,18	571,43	28,57	171	1,71
8	0,51	0,09	914,29	57,14	286	2,29
9	2,89	0,50	1010,00	50,00	106	1,00
10	1,202	0,26	666,67	38,10	187	1,52
11	1	0,18	628,57	57,14	246	2,86
12	3	0,53	666,67	57,14	145	0,19
13	12	2,00	420,00	40,00	71	0,45
14	2,4	0,53	142,86	26,67	192	1,14
15	0,9	0,17	446,16	47,59	291	3,57
15	1	0,18	285,71	57,14	314	2,86
17	0,48	0,09	342,86	57,14	343	3,43
18	1,2	0,26	171,43	57,14	232	1,90
19	1,2	0,26	179,05	57,14	263	1,90
20	1	0,18	342,86	57,14	240	2,86
21	1	0,18	240,00	57,14	274	1,71
22	1,5	0,26	190,48	45,71	331	1,90
23	4	0,70	285,71	57,14	213	1,43
24	1,5	0,26	190,48	45,71	259	1,90
25	1,68	0,38	289,47	39,47	200	2,63
26	0,55	0,10	700,00	50,00	210	1,00
27	2	0,35	1142,86	42,86	131	2,00
28	1,7	0,31	790,32	45,16	161	0,97
29	1	0,20	850,00	50,00	210	2,00
30	0,9	0,18	1388,89	66,67	311	3,61
⋮	⋮	⋮	⋮	⋮	⋮	⋮
689	1,5	0,30	666,67	26,67	353	0,67
690	0,8	0,13	447,76	29,85	254	0,75
691	1,7	0,30	833,33	26,67	433	1,00

## Lampiran 2. Syntax R Model Semiparametrik Campuran

```
options(digits=8)
data=read.csv('d:/datahesi.csv',header=TRUE,sep=";")
#data=data[1:10,]
SplineKernel=function(data,tknot,mx,mxker)
{
  #tknot=number knot point
  #mx=number of parametric
  #mxker=number of non parametric with kernel
  library(pracma)
  n=nrow(data)
  m=ncol(data)
  mnonpar=m-mx-mxker-1
  mk=mnonpar*tknot
  mker=mxker
  y=data[,1];
  x=as.matrix(cbind(1,data[,2:(mx+1)]))
  xnon=as.matrix(data[,((mx+2):(m-mxker))]);
  xker=as.matrix(data[, (m-mxker+1):m])
  if (tknot==1) nk=22 else if (tknot==2) nk=14 else nk=12
  m1.nn = matrix(1,nrow=n,ncol=n)
  satu = matrix(1,nrow=n,ncol=1)
  mi.nn = diag(1,nrow=n,ncol=n)
  knot=matrix(ncol=(mnonpar*tknot),nrow=nk)
  In=diag(n)
  for (i in 1:mnonpar)
    knot[,i]= seq(min(xnon[,i]),max(xnon[,i]),length.out=nk)
  knot=as.matrix(knot[2:(nk-1),])
  nk=nrow(knot)
  knot1=knot

  if (tknot==1){ # build knot point
    knot1=knot1;
  }else if (tknot==2){
    z=(nk*(nk-1)/2);
    knot2=matrix(nrow=(z+1),ncol=mk)
    for (i in 1:mnonpar)
      {
        knot=matrix(nrow=1,ncol=2)
        for (j in 1:(nk-1))
          {
            for ( k in (j+1):(nk))
              {
```



```

        xx=cbind(knot1[j,i], knot1[k,i])
        knot=rbind(knot,xx)
    }
}
knot2[, (i*2-1):(i*2)]=knot;
}
knot2=knot2[2:z+1,]
knot1=knot2
}else {
    z=(nk*(nk-1)*(nk-2)/6)
    knot2=matrix(nrow=(z+1),ncol=mk)
    for (i in 1:mnonpar)
    {
        knot=matrix(nrow=1,ncol=3)
        for (j in 1:(nk-2))
        {
            for ( k in (j+1):(nk-1))
            {
                for (g in (k+1):nk)
                {
                    xx=cbind(knot1[j,i] ,knot1[k,i] ,knot1[g,i])
                    knot=rbind(knot, xx)
                }
            }
        }
        knot2[, (i*3-2):(i*3)]=knot
    }
    knot2=knot2[2:z+1,]
    knot1=knot2;
}
knot=knot1
nk=nrow(knot)
matGCV=matrix(ncol=1,nrow=nk)
matMSE=matrix(ncol=1,nrow=nk)
matRsqr=matrix(ncol=1,nrow=nk)
matalfa=matrix(ncol=mxker,nrow=nk)
beta=matrix(ncol=(1+mx),nrow=nk)
teta=matrix(ncol=(mnonpar+(mnonpar*tknot)),nrow=nk)
nbetateta=(1+mx)+(mnonpar+(mnonpar*tknot))
Ix=diag((1+mx))
Ixnon=diag((mnonpar+(mnonpar*tknot)))
for(i in 1:nk)
{

```

```

cat('loading...',i/nk*100,"% ", "\n")
alfa=c(rep(0.05,mker))
matknot=matrix(ncol=(mnonpar*tknot),nrow=n)
GCVmin = function(alfa)
{
  sumWZ = 0
  for(j in 1:mker)
  {
    xkerdiag = diag(xker[,j])
    V = m1.nn%%xkerdiag
    Z = (t(V)-V)/alfa[j]
    K = 1/sqrt(2*pi)*exp(-1/2*Z^2) #fungsi gaussian
    KA = (1/alfa[j])*K
    Wbawah = diag(c(1/n*KA%%satu))%%m1.nn
    WZ = 1/n*KA/Wbawah
    sumWZ = sumWZ+WZ
  }
  Xker = sumWZ/mker

  for ( k in 1:(tknot*mnonpar))
  {
    if (tknot==1) { b=k;
    } else if (tknot==2) { if (mod(k,2)==1) b=floor(k/2)+1 else b=k/2
    } else {b=ceil(k/3);}
    matknot[,k]=pmax(0,(xnon[,b]-knot[i,k]));
  }
  Xnon=as.matrix(cbind(xnon,matknot))

  invXnon=pinv(t(Xnon)%%Xnon)%%t(Xnon)
  invx=pinv(t(x)%%x)%%t(x)

  B1=pinv(Ix-(invx%%Xnon%%invXnon%%x))
  B2=invx%%(In-Xker)-invx%%Xnon%%invXnon%%(In-Xker)
  Beta1=B1%%B2

  T1=pinv(Ixnon-(invXnon%%x%%invx%%Xnon))
  T2=invXnon%%(In-Xker)-invXnon%%x%%invx%%(In-Xker)
  Teta1=T1%%T2

  yhat = (x%%Beta1+Xnon%%Teta1+Xker)%%y
  SSE = sum((y-yhat)^2)
  MSE = SSE/n
  GCV = MSE/((n^1*sum(diag(mi.nn-

```

```

(x%%Beta1+Xnon%%Teta1+Xker)))^2))
  return(GCV)
}
optimasi = suppressWarnings(nlminb(c(rep(0.05,mker)),GCVmin,lower=0))
matalfa[i,] = optimasi$par
matGCV[i] = optimasi$objective
alfa = matalfa[i,]

sumWZ = 0
for(j in 1:mker)
{
  xkerdiag = diag(xker[,j])
  V = m1.nn%%xkerdiag
  Z = (t(V)-V)/alfa[j]
  K = 1/sqrt(2*pi)*exp(-1/2*Z^2) #fungsi gaussian
  KA = (1/alfa[j])*K
  Wbawah = diag(c(1/n*KA%%satu))%%m1.nn
  WZ = 1/n*KA/Wbawah
  sumWZ = sumWZ+WZ
}
Xker = sumWZ/mker
for ( k in 1:(tknot*mnonpar))
{
  if (tknot==1) { b=k;
  } else if (tknot==2) { if (mod(k,2)==1) b=floor(k/2)+1 else b=k/2
  } else {b=ceil(k/3);}
  matknot[,k]=pmax(0,(xnon[,b]-knot[i,k]));
}
Xnon=as.matrix(cbind(xnon,matknot))
invXnon=pinv(t(Xnon)%%Xnon)%%t(Xnon)
invx=pinv(t(x)%%x)%%t(x)

B1=pinv(Ix-(invx%%Xnon%%invXnon%%x))
B2=invx%%(In-Xker)-invx%%Xnon%%invXnon%%(In-Xker)
Beta1=B1%%B2

T1=pinv(Ixnon-(invXnon%%x%%invx%%Xnon))
T2=invXnon%%(In-Xker)-invXnon%%x%%invx%%(In-Xker)
Teta1=T1%%T2

yhat = (x%%Beta1+Xnon%%Teta1+Xker)%%y
beta[i,]=Beta1%%y
teta[i,]=Teta1%%y

```

```

SSE = sum((y-yhat)^2)
SSR=sum((yhat-mean(y))^2)
SST=SSR+SSE
matRsqr[i]=(SSR/(SST))*100
matMSE[i] = SSE/n
}
dataprint = data.frame(cbind(matGCV,matRsqr,matMSE,beta,teta,knot,matalfa))
mbeta=ncol(beta)
mteta=ncol(teta)
mknot=ncol(knot)
malfa=ncol(matalfa)
a<-c("GCV", "Rsqr", "MSE",paste("beta",c(0:(mbeta-
1))),paste("teta",c(1:(mteta))),paste("knot",c(1:(mknot))),paste("alfa",c(1:(malfa))))
colnames(dataprint)<-a
dataprint = dataprint[order(dataprint$GCV,dataprint$MSE),]
cat("\n", "Result : ", "\n")
nprint=min(20,nrow(dataprint))
print(dataprint[1:nprint,])
write.csv(dataprint,paste('d:/result estimation parameter knot ', tknot,'.csv'))
list(hasil=dataprint)
}

```

## BIOGRAFI PENULIS



**Hesikumalasari**, Lahir di Bima, NTB pada tanggal 26 Mei 1992. Penulis merupakan anak pertama dari empat bersaudara dari pasangan H. Sadam Ismail dan Hj. Rugayah M. Saleh. Penulis mulai memasuki jenjang pendidikan formal di Sekolah Dasar Impres Sangiang pada tahun 1998-2004. Kemudian penulis melanjutkan pendidikan di Sekolah Menengah Pertama Negeri 1 Wera tahun 2004-2007. Pada tahun yang sama penulis melanjutkan pendidikan ke Madrasah Aliyah Negeri 1 Kota Bima pada tahun 2007-2010. Penulis kemudian melanjutkan pendidikan Sarjana (S1) di Universitas Negeri Makassar (UNM) pada Program Studi Matematika dan selesai pada tahun 2014. Kemudian pada tahun yang sama Penulis melanjutkan studi Pasca Sarjana (S2) di Institut Teknologi Sepuluh Nopember Program studi Statistika. Segala kritik dan Saran yang dapat dikirim melalui email : [hesikumala@gmail.com](mailto:hesikumala@gmail.com).